

1/5/6 (Item 6 from file: 351)  
DIALOG(R) File 351:Derwent WPI  
(c) 2006 The Thomson Corp. All rts. reserv.

013857376 \*\*Image available\*\*  
WPI Acc No: 2001-341589/ 200136  
XRPX Acc No: N01-247411

Language translation apparatus has speech recognition input, prosody  
information extractor and speech output

Patent Assignee: SONY CORP (SONY )  
Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
JP 2001117922	A	20010427	JP 99293875	A	19991015	200136 B

Priority Applications (No Type Date): JP 99293875 A 19991015

Patent Details:

Patent No	Kind	Lan Pg	Main IPC	Filing Notes
JP 2001117922	A	32	G06F-017/28	

Abstract (Basic): JP 2001117922 A

NOVELTY - A speech recognition unit recognizes voice input into input unit. The recognized speech result is transmitted to machine translator (1) which is provided with conversion table. Based on conversion table which uses prosody information for one or both of the languages, the recognized speech result is translated into translation sentence and supplied to speech synthesizer for performing speech synthesis. The processed speech is output by composite tone from speaker (5).

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

(a) Translation procedure;

(b) Program for executing translation of language

USE - For translating Japanese to English and vice versa.

ADVANTAGE - Accurate translation is performed and translated sentence is output by speech synthesis, without imparting a feeling of incongruity.

DESCRIPTION OF DRAWING(S) - The figure shows block diagram of audio translation system. (Drawing includes non-English language text).

Machine translator (1)

Speaker (5)

pp; 32 DwgNo 2/24

Title Terms: LANGUAGE; TRANSLATION; APPARATUS; SPEECH; RECOGNISE; INPUT;  
PROSODY; INFORMATION; EXTRACT; SPEECH; OUTPUT

Derwent Class: P86; T01

International Patent Class (Main): G06F-017/28

International Patent Class (Additional): G06F-003/16; G10L-013/00;  
G10L-015/00

File Segment: EPI; EngPI

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-117922

(43)Date of publication of application : 27.04.2001

(51)Int.Cl.

G06F 17/28

G06F 3/16

G10L 13/00

G10L 15/00

(21)Application number : 11-293875

(71)Applicant : SONY CORP

(22)Date of filing : 15.10.1999

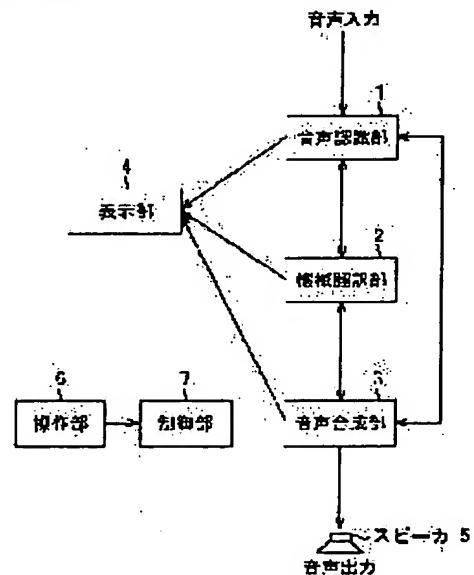
(72)Inventor : HIROE ATSUO  
TSUTSUMI HIRONAGA  
KISHI HIDEKI  
ASANO KOJI

## (54) DEVICE AND METHOD FOR TRANSLATION AND RECORDING MEDIUM

## (57)Abstract:

PROBLEM TO BE SOLVED: To perform high-accuracy translation and to output a translated sentence in synthetic voice without feelings of incompatibility.

SOLUTION: In a voice recognizing part 1, inputted voices are recognized, and the prosody information of inputted voices is extracted and supplied to a machine translation part 2 together with the voice recognized result of inputted voices. The machine translation part 2 has a translation table, in which language translation data for performing Japanese/English and English/Japanese translations are described together with prosody information concerning English and Japanese, and on the basis of that translation table, the voice recognized result is translated to a translated sentence with prosody information. This translated sentence with prosody information is supplied to a voice synthesizing part 3 and outputted in synthetic voices from a speaker 5 by performing ruled voice synthesizing processing.



音声翻訳システム

## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]



er took this picture. "brother"はアクセント付きで発音される。

【0012】一方、日本国による英語「弟は、この写真  
を撮りました。」については、「この写真」が新情報と  
なっており、「弟」が旧情報になっている。この日本語  
による英語の英訳は、「My brother took this picture  
a.」となり、新情報「この写真」の対訳「this picture」  
が、旧情報「弟」の対訳「My brother」よりも先行してい  
ないから、英訳「My brother took this picture. は、  
10 アクセントなしで英語される。

【0013】日本語による英語「この写真は、弟が撮りました。」と、「彼は、この写真を撮りました。」の英訳は、それぞれ上では、いずれも「My brother took this picture.」となり、同一であるが、上述のように、発話したときのアクセントが異なる。従って「アクセントを考慮せず」に、英英や英日の翻訳を行うと、以下のような問題が生じる。

【0014】即ち、'日英の照会を行い、その翻訳結果を、台成音で出力する場合、'日本語による発語「この写真は、弟が撮りました。」の英訳'brother+took+his picture、'台成音には、'brother'にアクセントがあるべきであり、また、'日本語による発語「弟は、この写真を撮りました。」の英訳'brother took this picture、'台成音には、'brother'にアクセントがあるべきでない。

【0016】しかしながら、アクセントを考慮しない場合は、いずれの英訳もアクセントなしの合成音で出力される。即ち、日本語による英訳「この写真は、弟が撮りました。」と、「弟は、この写真を撮りました。」の翻訳結果は、区別せずに、いずれも、"My brother took this picture."というアクセントなしの合成音で出力されることになる。従って、英語のユーザからすれば、日本語による英訳「この写真は、弟が撮りました。」の英訳として、意味的に違和感のある合成音が出力されることになる。

【0016】逆に、英日の翻訳を行う場合には、英語による英語“my brother's took this picture.”と、“My brother took this picture.”の翻訳結果は、上述のように、それぞれ「この写真は、弟が撮りました。」と、「弟は、この写真を撮りました。」となつて区別されるべきである。しかしながら、アクセントを考慮しない場合には、上述の英語による2つの英語は区別されず、い

40

ずれの日本語版も、「弟は、この写真を持ちました。」となり、日本語のユーザーからすれば、英語による発音「brother」が「brother」の日本語版として、意味的に違和感のある翻訳文が出力されることになる。【0017】また、例えば、日本語版は、朗句「は」と「弟」が使いつまられることがあり、さらに、朗句「弟」が使いつまられる朗句文が出力されることになる。【0018】「弟」が、その直前に配置される名前がアクセント付きで強調されることがあるが、アクセントを考慮して強調されることあるが、アクセントを考慮

しないで行くと、違和感のある翻訳結果が得られることがある。

【0018】具体的には、例えば、「ヘンリーがどうし  
たのだのか」という質問に対しての英語によるユーザの返答  
が、「Henry has arrived.」または「Henry has arrived  
.」であった場合においては、その日本語訳は、「ヘン  
リーは到着しました。」となる。

【0019】また、例えば、「誰が到着したのか」また  
は「到着したのは誰か」という質問に対しての英語によ  
るユーザの返答が、「Henry has arrived.」であった場  
合においては、その日本語訳は、「ヘンリーが到着し  
ました。」または「到着したのはヘンリーです。」とな  
る。

【0020】さらに、例えば、「何が起こったのか」という質問に対しての英語によるユーザの返答が、“Henry has arrived.”であった場合には、その日本語訳は、「ヘンリーが到着しました。」となる。

【0021】従って、日英の翻訳においては、「ヘンリー」は到着をした」、「ヘンリー」が到着しました、「ヘンリー」到着したのはヘンリーです」、「ヘンリー」が到着しました」の翻訳結果は、アクセントを考慮しないと、いずれも、"Henry has arrived."となり、この翻訳結果を、命令音で出力した場合には、英語のユーザにとって、意味や文法和感のある合成音が出力されることとなる。

【0022】一方、英日間の翻訳においては、“Henry has arrived.” “Henry<sup>が</sup> has arrived.” “Henry has <sup>が</sup> arrived.” の翻訳結果は、アクセントを考慮しないと、いづれも、「ヘンリーがどくたのか」という質問に対しての返答である「ヘンリーは到着した。」か、またでは、「何が起ったのか」という質問に対しての返答である「ヘンリーが到着しました。」となり、英語による英語におけるアクセントの違いによって、訳し分けを行うことができない。その結果、日本語のユーザにとつて、覚えるに違和感のある翻訳文が出力されることがあ

【0023】さらに、例えば、英語では、副詞と、その副詞の意味上の係り先との位置関係によって、アクセント付きの発話が行われることがあるが、アクセントを考慮しない副詞を行うと、違和感のある副詞結果が得られることがある。

【0 0 2 4】即ち、英語では、例えば、“also”や“only”といった副詞は、本能的には、それが修飾する語句の直前に配置されるが（“only”は、それが修飾する語句の直後に配置されることもある）、話し言葉では、これらの副詞が、動詞の直前に配置され、保り先（修飾先）の語句に、アクセント（ストレス）が付されることがある。

【0026】このため、例えば、英語による発語'I also  
o like there.'の日本語訳は、「私は彼女も好きで  
す。」になるべきであり、英語による発語'eis also li

インターネット・ネットを付加することで、その翻訳文に対応する合成音が生成されて出力される。なお、機械翻訳部202が出力する翻訳文は、図示せぬディスプレイで表示されることもある。

【0008】  
 (説明が情状しようとする語句) 音声組織システムにおいて、原音節による音声が、目的音節による正しい韻文文法に対峙する合成音で出力されるかどうかの処理状況に現れては、音声認識部201、組織制御部202、音声合成部203それぞれにおける処理の精度が問題となる。音声(原音節による音声)が、一旦、テキストに変換され、再び、音声(目的音節による合成音)に変換されるという処理過程で、情報が欠落することにも問題となる。

【０００７】即ち、音声には、読み取り可能な文字で表すことのできる音価情報（他の、アクセントや、イントネーション、リズム、ポーズといった韻律に関するプロンティ（prosody）情報も含まれるが、音声認識部１）において音声認識が行われ、その音声認識結果がテキストで出力されることにより、音声に含まれていたプロンティ情報は失われる。

「0008」しかしながら、文字によって表現した場合とは同一になる音声であっても、その意味内容が、音声に含まれるプロンディ情報によって異なる場合があり、上述のように、類似処理の過程で、キャストを介することによりプロンディ情報が欠落することが、類似結果に影響を与えることがある。

【0009】即ち、例へば、いまは、日本語と英語の話し手（ユーザ）どうしが対話をを行うために、日英および英日の翻訳を行うとして、アクセントまたは上り調子のイントネーションがある語句を、アスタリスク（\*）で囲んで記述するのは、例えば、「\*I have a pen」で「I」を強く発音することを教示し、また、「I have a pen」では、「pen」を強く発音することを教示する。

【0010】例えば、文中に、新情報（対話の中に、新たに現れた情報）である語句と、旧情報（対話の中に既に現れている情報）である語句とがある場合においては、真の情報は旧情報よりも先行して現れるときには、英語では、一般に、新情報の語句は、アクセント付きて発話される。従って、アクセントを考慮しないで翻訳を行うと、違和感のある翻訳結果が得られる（図った翻訳結果が得られる）ことがある。

【0011】具体的に、例えば、「弟が写真をとった」という事実がある場合において、日本語による発話「この写真は、弟が撮りました。」については、「この写真」が旧情報となっており、「弟」が新情報になっている。この日本語による発話の英訳（和訳文）は、“My brother\* took this picture.”となり、新情報（弟）の対訳“My brother”が、旧情報「この写真」の対訳“this picture”よりも先行して現れるから、英訳“My brother”

情報が与えられるため、ユーザの個人性を反映した合成音  
を出力するのは困難である。

【0033】また、例えば、ユーザの性別によって、そ  
のユーザによる発話の韻律結果を変化させた場合にお  
いては、性別の識別は、音声からは、比較的に容易に行  
うことが可能であるが、音声が、その音声認識結果とし  
てのテキストに変換された後に、その識別を行うのは困  
難である。

【0034】そこで、従来の音声認識や音声合成の手法  
の中には、音声のブロードバンド情報を用いて処理を行うも  
のがある。

【0035】即ち、例えば、特開平8-50498号公  
報には、音韻情報だけでなく、イントネーションの情報  
(音韻学的情報)をも用い、例えば、「橋」および  
「著」等といった同音異アクセント語を区別して、音声  
認識を行う方法が示されている。

【0036】しかしながら、特開平8-50498号公  
報に開示されている方法では、同一の単語列ではある  
が、アクセントのある単語が異なるもの(例えば、上述  
の「an English teacher」と、「an English tea  
cher」など)を区別して認識することはできない。さら  
に、特開平8-50498号公報に開示されている方法  
は、あくまで、同音異アクセント語を区別して音声認識  
を行うことができるだけで、その音声認識の結果出力さ  
れるテキストを認識し、さらに、その認識結果に対応す  
る合成音を生産する場合に、ブロードバンド情報が考慮さ  
れるわけではないため、上述したような違和感のある韻律  
文や合成音が出力されることに対応することはできな  
い。

【0037】また、例えば、特開平6-332494号  
公報には、原音語による入力音声から、アクセントのあ  
る語句を抽出し、その語句に対応する目的語の語句に  
アクセントを付す韻律結果が示されている。

【0038】しかしながら、特開平6-332494号  
公報では、原音語においてアクセントのある語句に対応  
する目的語の語句に、アクセントがある場合(例え  
ば、上述の「Henry has arrived.」が「ヘンリーが到  
着しました。」と翻訳される場合や、「Henry has arriv  
ed.」が「ヘンリーが到着しました。」「と翻訳される場合  
など)には対応することができず、原音語においてア  
クセントのある語句に対応する目的語の語句に、アク  
セントがない場合(例えば、上述の「Henry has arrive  
d.」が「ヘンリーは到着しました。」「と翻訳される場合  
など)において原音語においてアクセントのない語句に  
対応する目的語の語句に、アクセントがある場合に対  
応するのは困難である。

【0039】本発明は、このような状況に鑑みてなされ  
たものであり、精度の高い翻訳を行うことや、翻訳文  
を、違和感のない合成音で出力することができるよう  
するものである。

【0040】

【課題を解決するための手段】本発明の翻訳装置は、入  
力文を入力する入力手段と、入力文を、翻訳文に翻訳す  
るための対応関係が、第1または第2の言語のうちの少  
なくとも一方のブロードバンド情報とともに記述されてい  
るテーブルに基づいて、入力文を、その入力文に対応する  
翻訳文に翻訳する翻訳手段と、翻訳文を出力する出力手  
段とを備えることを特徴とする。

【0041】入力手段には、音声認識する音声認識手  
段を設け、音声認識手段による音声認識結果を、入力文  
として出力することができる。

【0042】音声認識手段には、音声のブロードバンド情  
報を抽出する抽出手段を設け、音声の音声認識結果を、そ  
の音声のブロードバンド情報とともに出力させることがで  
きる。翻訳手段には、音声認識結果を、そのブロードバン  
ド情報を用いて翻訳させることができる。

【0043】本発明の翻訳装置には、音声認識手段に対  
して、ブロードバンド情報を要求する要求手段をさらに設け  
ることができ、この場合、音声認識手段には、要求手段  
による要求があった場合に、ブロードバンド情報を出力させ  
ることができ、

【0044】出力手段には、翻訳文に対応する合成音を  
生成する音声合成手段を設け、翻訳文を、合成音で出力  
させることができる。

【0045】翻訳手段には、翻訳文を、そのブロードバン  
ド情報とともに出力させ、音声合成手段には、翻訳文に対  
応する合成音を、そのブロードバンド情報を用いて生成さ  
せることができる。

【0046】音声合成手段には、翻訳手段に対して、ブ  
ロードバンド情報を要求する要求手段を設けることができ、  
この場合、翻訳手段には、要求手段による要求があった  
場合に、ブロードバンド情報を出力させることができる。

【0047】本発明の翻訳装置には、入力文を、翻訳文  
に翻訳するための対応関係が、第1または第2の言語の  
うちの少なくとも一方のブロードバンド情報とともに記述さ  
れているテーブルを記憶している記憶手段をさらに設け  
ることができる。

【0048】本発明の翻訳方法は、入力文を入力する入  
力ステップと、入力文を、翻訳文に翻訳するための対応  
関係が、第1または第2の言語のうちの少なくとも一方  
のブロードバンド情報とともに記述されているテーブルに基  
づいて、入力文を、その入力文に対応する翻訳文に翻訳  
する翻訳ステップと、翻訳文を出力する出力ステップと  
を備えることを特徴とする。

【0049】本発明の翻訳装置は、入力文を入力する入  
力ステップと、入力文を、翻訳文に翻訳するための対応  
関係が、第1または第2の言語のうちの少なくとも一方  
のブロードバンド情報とともに記述されているテーブルに基  
づいて、入力文を、その入力文に対応する翻訳文に翻訳  
する翻訳ステップと、翻訳文を出力する出力ステップと  
を備えることを特徴とする。

を備えるプログラムが記憶されていることを特徴とす  
る。

【0050】本発明の翻訳装置および翻訳方法、並びに  
記録媒体においては、入力文を、翻訳文に翻訳するため  
の対応関係が、第1または第2の言語のうちの少なくと  
も一方のブロードバンド情報とともに記述されているテー  
ブルに基づいて、入力文が、その入力文に対応する翻訳文  
に翻訳される。

【0051】

【発明の実施の形態】図2は、本発明で適用した音声認  
知システム(システム)とは、複数の装置が協力的に集合  
したものをいい、各構成の装置が同一筐体中にあるか否か  
は問わない。一実施の形態の電気的情報例を示してい  
る。図3は、その音声認識システムの外部構成例を示し  
ている。

【0052】この音声認識システムでは、日本語や英語  
等の原音語による音声が入力されると、その音声を、英  
語や日本語等の目的語に翻訳した翻訳文が出力される  
ようになっている。

【0053】即ち、例えば、日本語や英語などによる音  
声は、マイク11に入力され、音声認識部1に供給され  
る。音声認識部1は、マイク11からの音声信号を音声認識  
し、その音声認識結果としてのテキスト、その他付随す  
る情報を、機械翻訳部2や、表示部4などに出力する。

【0054】機械翻訳部2は、音声認識部1が出力する  
音声認識結果を解析し、入力された音声の言語(原音  
語)を、例えば英語や日本語などの目的語に機械翻訳  
し、その翻訳結果としてのテキスト、その他付随する情  
報を、音声合成部3や、表示部4などに出力する。音声  
合成部3は、機械翻訳部2などの出力に基づいて音声合  
成処理を行い、入力された音声の、他の言語への翻訳結  
果等としての合成音を出力する。

【0055】表示部4は、例えば、液晶ディスプレイ等  
で構成され、音声認識部1による音声認識結果や、機械  
翻訳部2による機械翻訳結果等を表示する。

【0056】操作部6は、例えば、カーソルを移動させ  
る場合等に操作されるカーソルキー6A、選択を決定す  
る場合等に操作される決定キー6B、および選択をキャ  
ンセルする操作等に操作されるキャンセルキー6Cで構  
成される。操作部6の操作に対応する操作信号は、制御部  
7に供給されるようになっている。制御部7は、操作部  
6からの操作信号にしたがって、各種の処理を行うよう  
になっている。なお、操作部6は、上述した場合の他、  
文字の入力や、仮名漢字変換を行うとき等にも用いるこ  
とができるようになっている。

【0057】以上のように構成される音声認識システム  
においては、原音語による音声が入力されると、その音  
声は、音声認識部1で音声認識され、機械翻訳部2に供  
給される。機械翻訳部2では、音声認識部1による音声  
認識結果が、目的語に機械翻訳され、音声合成部3に

供給される。音声合成部3では、機械翻訳部2からの翻  
訳結果に対応する合成音が生成されて出力される。

【0058】次に、図4は、図2の音声認識部1の構成  
例を示している。

【0059】ユーザによる原音語の発話は、マイク11  
に入力され、マイク11では、その発話が、電気信号と  
しての音声信号に変換される。この音声信号は、AD(A  
nalog Digital)変換部12に供給される。AD変換部1  
2では、マイク11からのアナログ信号である音声信号  
がサンプリング、量子化され、デジタル信号である音  
声データに変換される。この音声データは、特徴抽出  
部13およびバッファ部14に供給される。

【0060】特徴抽出部13は、AD変換部12からの  
音声データについて、適当なフレームごとに、例えば、  
スペクトルや、線形予測係数、ケプストラム係数、線  
スペクトル対等の特徴パラメータを抽出し、バッファ部1  
4およびマッピング部15に供給する。

【0061】マッピング部15は、特徴抽出部13から  
の特徴パラメータに基づき、音響モデルデータベース1  
6、辞書データベース17、および文法データベース1  
8を必要に応じて参照しながら、マイク11に入力され  
た音声(入力音声)を認識する。

【0062】即ち、音響モデルデータベース16は、音  
声認識する音声の言語における個々の音素や音節などの  
音響的な特徴を表す音響モデルを記憶している。ここ  
で、音響モデルとしては、例えば、HMM(hidden Mark  
ov Model)などを用いることができる。辞書データベー  
ス17は、認識対象の各単語(語句)について、その発  
音に関する情報が記述された単語辞書を記憶している。こ  
で、文法規則としては、例えば、文法自由文法(CF

C)や、統計的な単語連動確率(N-gram)などに  
基づく規則を用いることができる。

【0063】マッピング部15は、辞書データベース1  
7の単語辞書を参照することにより、音響モデルデー  
タベース16に記憶されている音響モデルを選択すること  
で、単語の音響モデル(単語モデル)を構成する。さら  
に、マッピング部15は、個々の単語モデルを、文法  
データベース18に記憶された文法規則を参照すること  
により選択し、そのようにして接続された単語モデルを  
用いて、特徴パラメータに基づき、例えば、HMM法等  
によって、マイク11に入力された音声認識する。

【0064】そして、マッピング部15による音声認識  
結果は、例えば、原音語によるテキストで出力される。

【0065】ここで、音響モデルデータベース16に  
は、汎用的な音響モデルの他、必要に応じて、男性の  
声や、女性の声、年齢的な音質(例えば、笑い、  
怒り、くしゃみ、舌打ち等)等の、いわゆる特殊な音響モ

でも記憶させておくことができる。この場合、マッチング部 1.15 において、汎用的な音響モデルだけでなく、特殊な音響モデルをもつて音声認識を行うことで、各音響モデルによる音声認識結果の尤度に基づき、入力された音声か、男性の発話であるか、または女性の発話であるか、音打ち、あるいは笑い声であるかを判定することが可能となる。

[0066] なお、マッチング部 1.15 では、通常は、汎用的な音響モデルだけを音響認識を行い、機械的部 2 や音声合成部 3 から、特殊な音響モデルを用いて音声認識を行うように要求するリクエスト番号を用いて音声認識した場合には、特殊な音響モデルをもつて音声認識を行うことができる。

[0067] 一方、パッチング部 1.14 は、音声データバッファ 1.4A および特徴量バッファ 1.4B で構成され、音声データバッファ 1.4A は、AD 変換部 1.2 が出力する音声データを、特徴量バッファ 1.4B は、特徴抽出部 1.3 が出力する特徴パラメータを、それぞれ一時記憶する。

[0068] そして、音声データバッファ 1.4A に記憶された音声データや、特徴量バッファ 1.4B に記憶された特徴パラメータは、必要に応じて、プロンディ情報抽出部 1.9 によって読み出され、プロンディ情報抽出部 1.9 は、その音声データや特徴パラメータを用いて、入力された音声（入力音声）のプロンディ情報を抽出し、そのプロンディ情報をプロンディデータとして、マッチング部 1.15 が出力する音声認識結果としてのテキストに付随する情報として出力する。

[0069] なお、プロンディ情報抽出部 1.9 には、常に、プロンディ情報を抽出させて、プロンディデータを出力させることも可能であるし、機械的部 2 や音声合成部 3 から、プロンディ情報を要求するリクエスト番号を受信した場合にのみ、プロンディ情報を抽出させるようにすることも可能である。

[0070] また、プロンディ情報抽出部 1.9 が出力するプロンディデータは、マッチング部 1.15 に出力するようになっている。この場合、マッチング部 1.15 は、音声認識結果としてのテキストの中に、プロンディ情報抽出部 1.9 からプロンディデータを含まれた形のデータを生成して出力する。

[0071] 次に、図 5 は、図 2 の機械的部 2 の構成例を示している。

[0072] テキスト解析部 2.1 には、音声認識部 1 が出力する音声認識結果としてのテキストが、機械的部 2 の対象として入力されるようになっている。テキスト解析部 2.1 は、音声データ 2.3 や解析用文法データ 2.4 を用いて、音声データ 2.3 における各単語（語句）の表記を、解析用文法を適用するするために必要な品詞情報などを記憶された単語辞書が記憶されている。

また、解析用文法データ 2.5 には、単語辞書に記述された各単語の形態に基づいて、単語連綴に関する制約等が記述された解析用文法規則が記憶されている。そして、テキスト解析部 2.1 は、その単語辞書や解析用文法規則に基づいて、そこに入力されるテキスト（入力テキスト）の形態素解析や、構文解析等の解析情報を抽出する。ここで、テキスト解析部 2.1 における解析方法としては、例えば、正規文法や、文脈自由文法、統計的な単語連綴確率を用いたものなどがある。

[0074] テキスト解析部 2.1 で得られた入力テキストの解析結果としての単語情報は、音声認識部 2.2 に供給される。音声認識部 2.2 は、音声データ 2.2 を参照し、入力テキストの単語（原単語）の単語情報を、解析結果の単語（目的単語）の単語情報に変換する。

[0075] 即ち、音声認識部 2.2 には、原単語（原単語）の単語情報と、目的単語（目的単語）の単語情報と、音声認識部 2.2 から出力の単語の単語情報への変換パターンや、原単語と目的単語との対応関係例としてその対応関係と原単語との類似度の計算に用いられるシノニム等の、単語情報を記憶するための、原単語と目的単語との対応関係を記憶した単語データベースが記憶されている。そして、音声認識部 2.2 では、このような単語データベースに基づいて、入力テキストの単語の単語情報、目的単語の単語情報に変換される。

[0076] 音声認識部 2.2 で得られた目的単語の単語情報は、テキスト生成部 2.3 に供給される。テキスト生成部 2.3 は、音声データ 2.3 および生成用文法データ 2.8 を参照することにより、音声認識部 2.2 から目的単語の単語情報から、入力テキストを目的単語に翻訳したテキストを生成する。

[0077] 即ち、音声データ 2.7 には、目的単語の文を生成するのに必要な単語（語句）の品詞や活用等の情報が記述された単語辞書が記憶されており、また、生成用文法データ 2.8 には、目的単語の文を生成するのに必要な単語の活用規則や品詞の制約等の生成用文法規則が記憶されている。そして、テキスト生成部 2.3 は、これらの単語辞書および生成用文法規則に基づいて、音声認識部 2.2 から単語情報を、テキストに変換して出力する。

[0078] 次に、図 6 (A) の変換テーブルでは、原単語によるテキストとしては、いずれも「English teacher」と表現される目的単語によるテキスト「イギリス人の先生」および「英語の先生」の対応が、プロンディ情報としての単語と目的単語の対応が、プロンディ情報として、日本語「英語の先生」が対応付けられている。また、図 6 (B) では、英語「English teacher」の対応が、プロンディ情報として、日本語「イギリス人の先生」が対応付けられている。

[0084] この場合、音声認識部 2.2 は、変換テーブルを参照することで、英語による発話「English teacher」を、その音声認識結果とプロンディデータを用いて、日本語「イギリス人の先生」に正確に翻訳することができる。また、音声認識部 2.2 は、英語による発話「English teacher」を、その音声認識結果とプロンディデータを用いて、日本語「英語の先生」に正確に翻訳することができる。

[0086] 図 6 (B) の変換テーブルでは、英語「NP VP」の対応として、日本語「NP」が対応付けられており、英語「NP VP」の対応として、日本語「NP」が対応付けられている。ここで、NP は、名詞句 (Noun Phrase) を、VP は、動詞句 (Verb Phrase) を、それぞれ表

す。この場合、音声認識部 2.2 は、英語による発話「NP VP」を、その音声認識結果とプロンディデータを用いて、日本語「NP VP」に正確に翻訳することができる。また、英語による発話「NP VP」も、日本語「NP VP」に正確に翻訳することができる。

[0086] 図 6 (C) の変換テーブルでは、日本語「NP VP」および「NP VP」の対応として、いずれも、英語「NP VP」が対応付けられており、日本語「NP VP」の対応として、英語「NP VP」が対応付けられている。この場合、音声認識部 2.2 では、日本語「NP VP」の対応として、英語「NP VP」が出力され、音声合成部 3 では、その英語「NP VP」に対応する合成音、即ち、アクセントなしの合成音が生じられる。また、音声認識部 2.2 では、日本語「NP VP」の対応として、英語「NP VP」が出力され、音声合成部 3 では、その英語「NP VP」に対応する合成音、即ち、アクセントを付加した合成音が生じられる。従って、結果として、違和感のない合成音が生じられる。

[0087] 図 6 (D) の変換テーブルでは、英語「NP VP」の対応として、日本語「NP VP」が出力され、音声合成部 3 では、その英語「NP VP」に対応する合成音、即ち、名詞句 NP にアクセントを付加した合成音が生じられる。従って、結果として、違和感のない合成音が生じられる。

[0088] なお、図 6 の変換テーブルでは、強調を表す（強く発音される語句を表す）のに、アスタリスクを用いるようにしたが、強調その他のプロンディ情報を表す際には、例えば、HTML (Hyper Text Markup Language) 等で採用されているタグを用いることも可能である。

図 6 の変換テーブルのうち、例えば、アクセントだけに注目した場合でも、音節によって、高低型や強調型が存在し、そのような音節を区別したタグによってプロンディ情報を表すことは可能である。そこで、変換テーブルの作成の際には、そのような音節を区別しないアスタリスク等の記号を用いて、いわば簡略化された変換テーブルの記述を行い、後で、変換ツール等を用いて、簡略化された変換テーブルを、タグを用いて記述された変換テーブルにコンパイルするようにすることができ

[0089] 次に、図 7 は、図 2 の音声合成部 3 の構成例を示している。

[0090] テキスト解析部 3.1 には、機械的部 2 が出力する音声認識結果としてのテキストが、音声合成部 3 の対象として入力されるようになっている。テキスト解析部 3.1 は、音声データ 3.3 や解析用文法データ 3.4 を参照しながら、そのテキストを解析する。

でも記憶させておくことができる。この場合、マッピング部15において、汎用的な音響モデルだけでなく、特殊な音響モデルをも利用して音声認識を行うことで、各音響モデルによる音声認識結果の尤度に基づき、入力された音声か、男性の発語であるか、または女性の発語であるかや、音質、あるいは笑い声であるかを判定することができる。

【0066】なお、マッピング部15では、通常は、汎用的な音響モデルだけを用いて音声認識を行い、機械部22や音声合成部23から、特殊な音響モデルを用いた場合にのみ、特殊な音響モデルをも利用して音声認識を行うようにすることが可能である。

【0067】一方、パツア部14は、音声データバッファ14Aおよび仲量バッファ14Bで構成され、音声データバッファ14Aは、AD変換部12が出力する音声データを、仲量バッファ14Bは、特徴抽出部13が出力する特徴パラメータを、それぞれ一時記憶する。

【0068】そして、音声データバッファ14Aに記憶された音声データや、仲量バッファ14Bに記憶された特徴パラメータは、必要に応じて、プロンディ情報抽出部19によって読み出され、プロンディ情報抽出部19は、その音声データや特徴パラメータを用いて、入力された音声（入力音声）のプロンディ情報を抽出し、そのプロンディ情報を表すプロンディデータを、マッピング部15が出力する音声認識結果としてのテキストに付随する情報として出力する。

【0069】なお、プロンディ情報抽出部19には、常に、プロンディ情報を抽出させて、プロンディデータを出力させることも可能である。機械部22や音声合成部23から、プロンディ情報を要求するリクエスト信号を受信した場合にのみ、プロンディ情報を抽出させるようにすることも可能である。

【0070】また、プロンディ情報抽出部19が出力するプロンディデータは、マッピング部15に出力するようになっている。この場合、マッピング部15は、音声認識結果としてのテキストの中に、プロンディ情報抽出部19からのプロンディデータを含めた形のデータを生成して出力する。

【0071】次に、図5は、図2の機械部22の構成例を示している。

【0072】テキスト解析部21には、音声認識部1が出力する音声認識結果としてのテキストが、機械部22として入力されるようになっている。テキスト解析部21は、辞書データベース24や解析用文法データベース25を参照しながら、そのテキストを解析する。【0073】即ち、辞書データベース24には、各単語（語句）の表記や、解析用文法を適用するために必要な品詞情報などが記述された単語辞書が記憶されている。

また、解析用文法データベース25には、単語辞書に記述された各単語の情報に基づいて、単語連綴に関する制約等が記述された解析用文法規則が記憶されている。そして、テキスト解析部21は、その単語辞書や解析用文法規則に基づいて、そこに入力されるテキスト（入力データ）の形態素解析や、構文解析等を行い、その入力テキストを構成する単語や構文の情報や単語情報を抽出する。ここで、テキスト解析部21における解析方法としては、例えば、正規文法や、文脈自由文法、統計的な単語連綴確率を用いたものなどがある。

【0074】テキスト解析部21で得られた入力テキストの解析結果としての単語情報は、言語変換部22に供給される。言語変換部22は、言語変換データベース26を参照し、入力テキストの単語（目的語）の単語情報を、翻訳結果の言語（目的語）の単語情報に変換する。

【0075】即ち、言語変換データベース26には、原言語（言語変換部22への入力の言語）の単語情報から、目的語（言語変換部22からの出力の言語）の単語情報への変換パターンや、原言語と目的語との対応関係およびその対応関係と原言語との間の類似度の計算に用いられるシソーラス等の、単語情報を記憶するための、原言語と目的語との対応関係を記述した言語変換データが記憶されている。そして、言語変換部22では、このような言語変換データに基づいて、入力テキストの単語の単語情報が、目的語の単語情報に変換される。

【0076】言語変換部22で得られた目的語の単語情報は、テキスト生成部23に供給される。テキスト生成部23は、辞書データベース27および生成用文法データベース28を参照することにより、言語変換部22からの目的語の単語情報から、入力テキストを目的語に翻訳したテキストを生成する。

【0077】即ち、辞書データベース27には、目的語の文を生成するのに必要な単語（語句）の品詞や活用等の情報が記述された単語辞書が記憶されており、また、生成用文法データベース28には、目的語の文を生成するのに必要な単語の活用規則や構文の制約等の生成用文法規則が記憶されている。そして、テキスト生成部23は、これらの単語辞書および生成用文法規則に基づいて、言語変換部22からの単語情報を、テキストに変換して出力する。

【0078】ここで、テキスト解析部21、言語変換部22、およびテキスト生成部23は、必要に応じて、音声認識部1が出力するプロンディデータを用いて処理を行うようになっている。テキスト解析部21や、言語変換部22、テキスト生成部23がプロンディデータを用いて処理を行った場合には、必要に応じてプロンディデータを含んだ処理結果が出力される。なお、音声認識部1が、常に、プロンディデータを出力するようになっ

いない場合には、テキスト解析部21、言語変換部22、およびテキスト生成部23は、プロンディ情報が必要ときに、それを要求するリクエスト信号を、音声認識部1に出力するようになっている。

【0079】また、テキスト解析部21には、処理中に参照した、辞書データベース24および解析用文法データベース25の情報を保持しておくことができる。同様に、言語変換部22や、テキスト生成部23にも、処理中に参照した、言語変換データベース26の情報や、辞書データベース27および生成用文法データベース28の情報を保持しておくことができる。この場合、テキスト解析部21や、言語変換部22、テキスト生成部23において、どの情報を参照して処理を行ったかの間い合わせのリクエスト信号が、後で処理を行うブロック（例えば、音声合成部3）から送情されてきたときに、そのブロックに対して、リクエスト信号が要求する情報（例えば、プロンディデータ等）を返信することが可能となる。

【0080】次に、図6は、図5の言語変換データベース26に記憶されている言語変換データを示している。【0081】本実施の形態では、言語変換データとして、原言語を目的語に翻訳するための対応関係が、原言語または目的語のうちの少なくとも一方のプロンディ情報とともに記述されており、そのような言語変換データが、変換テーブルに登録されている。

【0082】即ち、図6は、日本語または英語のうちの一方を、原言語とするともに、他方を、目的語とした場合の変換テーブルを示している。

【0083】図6（A）の変換テーブルでは、原言語によるテキストとしては、いずれも「English teacher」と表現される目的語によるテキスト「イギリス人の先生」および「英語の先生」の対応が、プロンディ情報としての加えられている部分を表す記号を用いて記述されている。即ち、図6（A）では、英語「English teacher」の対応として、日本語「イギリス人の先生」が対応付けられており、英語「English teacher」の対応として、日本語「英語の先生」が対応付けられている。

【0084】この場合、言語変換部22は、変換テーブルを参照することで、英語による英語「English teacher」を、その音声認識結果とプロンディデータを用いて、日本語「イギリス人の先生」に正確に翻訳することができ、また、言語変換部22は、英語による英語「English teacher」を、その音声認識結果とプロンディデータを用いて、日本語「英語の先生」に正確に翻訳することができる。

【0085】図6（B）の変換テーブルでは、英語「NP」の対応として、日本語「NPはVP」が対応付けられており、英語「NP+VP」の対応として、日本語「NPがVP」が対応付けられている。ここで、NPは、名詞句（Noun Phrase）を、VPは、動詞句（Verb Phrase）を、それぞれ表

す。この場合、言語変換部22は、英語による英語「NP+VP」を、その音声認識結果とプロンディデータを用いて、日本語「NPはVP」に正確に翻訳することができる。また、英語による英語「NP+VP」も、日本語「NPがVP」に正確に翻訳することができる。

【0086】図6（C）の変換テーブルでは、日本語「NPはVP」および「NPがVP」の対応として、いずれも、英語「NP+VP」が対応付けられており、日本語「NPがVP」の対応として、英語「NP+VP」が対応付けられている。この場合、言語変換部22では、日本語「NPはVP」や「NPがVP」の対応として、英語「NP+VP」が出力され、音声合成部3では、その英語「NP+VP」に対応する合成音、即ち、アクセントなしの合成音が生成される。また、言語変換部22では、日本語「NP+VP」の対応として、英語「NP+VP」が出力され、音声合成部3では、その英語「NP+VP」に対応する合成音、即ち、名詞句NPにアクセントを付加した合成音が生成される。従って、翻訳結果として、違和感のない合成音が出来上がる。

【0087】図6（D）の変換テーブルでは、英語「NP1 also V NP2」の対応として、日本語「NP1もNP2をV」が対応付けられており、英語「NP1+also V NP2」の対応として、日本語「NP1もNP2をV」が対応付けられている。ここで、Vは動詞を表す。この場合も、言語変換部22は、英語による英語「NP+NP」also V NP2を、日本語「NP1はNP2もV」に正確に翻訳することができ、また、英語による英語「NP1+also V NP2」を、日本語「NP1もNP2をV」に正確に翻訳することができる。

【0088】なお、図6の実施の形態では、強調を表す（強く発音される語句を表す）のに、アクセントを用いるようにしたが、強調その他のプロンディ情報を表すのには、例えば、HTML (Hyper Text Markup Language) 等で採用されているタグを用いることも可能である。

但し、プロンディ情報のうちの、例えば、アクセントだけに注目した場合でも、音質によって、高次元や低次元のソディ情報を記述することは可能である。そこで、変換テーブルの作成の際には、そのような型を区別しないアクセタリスク等の記号を用いて、いわば簡略化された変換テーブルの記述を行い、後で、変換ツール等を用いて、簡略化された変換テーブルを、タグを用いて記述された変換テーブルにコンバートするようにすることができ

る。

【0089】次に、図7は、図2の音声合成部3の構成例を示している。

【0090】テキスト解析部31には、機械部22が出力する翻訳結果としてのテキストが、音声合成部31の対象として入力されるようになっている。テキスト解析部31は、辞書データベース34や解析用文法データベース35を参照しながら、そのテキストを解析する。

【0091】即ち、辞書データベース34には、各単語(語句)の品詞情報や、読み、アクセント等の情報が記述された単語辞書が記憶されており、また、解析用文法データベース35には、辞書データベース34の単語辞書に記述された単語について、単語道順に関する制約等の解析用文法規則が記憶されている。そして、テキスト解析部31は、この単語辞書および解析用文法規則に基づいて、そこに入力されるテキストの形態素解析や構文解析等の解析を行い、後段の規則合成部32で行われる規則音声合成に必要な情報を抽出する。ここで、規則音声合成に必要な情報としては、例えば、ポーズの位置や、アクセントおよびイントネーションを制御するための情報その他の関係情報や、各単語の発音等の音韻情報などがある。

【0092】テキスト解析部31で得られた情報は、規則合成部32に供給され、規則合成部32では、音素片データベース36を用いて、テキスト解析部31に供給されたテキストに対応する合成音の音声データ(ディジタルデータ)が生成される。

【0093】即ち、音素片データベース36には、例えば、CV(Consonant, Vowel)や、VCV、CVC等の形で音素片データが記憶されており、規則合成部32は、テキスト解析部31から得られた情報に基づいて、必要な音素片データを探索し、さらに、ポーズ、アクセント、イントネーション等を適切に付加することで、テキスト解析部31に供給されたテキストに対応する合成音の音声データを生成する。

【0094】この音声データは、DA変換部33に供給され、そこで、アナログ信号としての音声信号に変換される。この音声信号は、指示されたスピーカに供給され、これにより、テキスト解析部31に供給されたテキストに対応する合成音が出力される。

【0095】ここで、テキスト解析部31、規則合成部32、およびDA変換部33は、必要に応じて、音声部15に出力する音声データや、機械翻訳部2が出力する翻訳結果に含まれるブロンディデータを用いて処理を行うようになっている。なお、音声部15が、常に、ブロンディデータを出力するようになっている場合には、テキスト解析部31、規則合成部32、およびDA変換部33は、ブロンディ情報が必要なときに、それを要求するリクエスト信号を、音声部15に出力するようになっている。同様に、機械翻訳部2が、常に、ブロンディを含む処理結果を出力するようになっている場合には、テキスト解析部31、規則合成部32、およびDA変換部33は、ブロンディ情報が必要なときに、それを要求するリクエスト信号を、機械翻訳部2に出力するようになっている。

【0096】なお、本実施の形態では、音声部15は、ブロンディ情報を表すタグ(以下、適宜、ブロンディタグという)を用いて記述されたテキストを処理する

区切りが、それぞれの開始時刻および終了時刻とともに記述され、さらに、高低アクセントおよび強弱アクセントについては、その度合いも記述されている。なお、高低アクセントや強弱アクセントの度合いは、0を基準として、高いまたは強い場合をプラスの数字で、低いまたは弱い場合をマイナスの数字で、それぞれ表してある。

【0112】図11(B)においては、ブロンディ情報としての高アクセントの開始時刻および終了時刻だけが記述されている。後段で処理を行う機械翻訳部2や音声部15が、上述したようなブロンディ情報のすべてを必要とせず、ある高さ以上の高アクセントの有無だけを必要としている場合には、ブロンディ情報抽出部19において、すべてのブロンディ情報を抽出するのは無駄であり、必要なもののみ抽出すればよい。さらに、必要なブロンディ情報は、上述のように、高アクセントのみであり、しかも、ある高さ以上の高アクセントの有無だけである場合には、図11(A)に示したように、ブロンディ情報の種類や、度合いは不要である。そこで、このような場合には、図11(B)に示したような、ある高さ以上の高アクセントというブロンディ情報が存在する期間時刻および終了時刻だけが記述されたブロンディデータを用いることが可能である。

【0113】なお、ブロンディデータは、次に説明するように、マッピング部15において、その音声認識結果に含まれるが、その際、ブロンディタグ付き認識結果、ブロンディタグの例を、図11(C)に示す。

【0114】再び、図8に戻り、マッピング部15は、ブロンディ情報抽出部19からブロンディデータを受信すると、ステップS3において、そのブロンディデータを、音声認識結果に含まれたブロンディタグ付き認識結果を生成し、機械翻訳部2に出力する。

【0115】機械翻訳部2(図5)では、ステップS4において、音声認識部1からのブロンディタグ付き認識結果を用いて、目的言語への機械翻訳が行われる。

【0116】即ち、機械翻訳部2では、テキスト解析部21において、音声認識部1からのブロンディタグ付き認識結果を用いて、原言語による発話の音声認識結果のテキスト解析が行われ、その解析結果が、言語変換部22に供給される。ここで、テキスト解析部21において、ブロンディタグを参照することで、上述したような翻訳結果が行われるような解析が行われることを防止することができる。

【0117】言語変換部22は、言語変換データを用いて、原言語のブロンディタグ付き認識結果することにより、原言語のブロンディタグ付き認識結果を翻訳し、目的言語のブロンディタグ付き認識結果に変換する。このブロンディタグ付き認識結果は、テキスト生成部23に供給され、そこで処理された後、音声部15に出力される。

【0118】音声部15(図7)では、ステップS6において、機械翻訳部2からのブロンディタグ付き翻訳

報として、例えば、プリミティブなもの、ユーザ(人)の発音を反映したユーザの個人性を反映したもの、あるいは抽象化したものが抽出されるようになっている。プリミティブなブロンディ情報としては、例えば、入力音声のピッチや、パワー、発話速度、発話時間、ポーズの長さ等がある。また、ユーザの個人性を反映したブロンディ情報、あるいは抽象化したブロンディ情報としては、強弱アクセント/イントネーションや、高低アクセント/イントネーション、男性または女性のいずれの声であるか、どのような感情の口調であるか(例えば、怒っている口調であるとか、笑っている口調であるとかなど)、非言語的な音情報(例えば、舌打ちや、くしゃみの情報)、平綴文であるか、または疑問文であるか、句の区切り位置等がある。

【0105】ブロンディ情報抽出部19では、プリミティブなブロンディ情報は、音声データや、その特徴パラメータを用いて計算を行うことで求められ、ユーザの個人性を反映したブロンディ情報、あるいは抽象化したブロンディ情報は、プリミティブなブロンディ情報の1つ以上を用いて求められる。

【0106】即ち、例えば、句の区切り位置は、その位置におけるポーズの長さや、高低イントネーションの変化を用いて求められる。具体的には、例えば、「全員ロッカー支給」という日本語は、そのテキストだけを見ても、「全員に、ロッカーを支給する」という意味にもとれし、「全員ロッカー」という句があつて、それを支給する」という意味にもとれる。この場合、「全員ロッカー支給」という日本語が、例えば、「全員に、ロッカーを支給する」という意味であるとする、「全員ロッカー支給」という日本語の「全員」と「ロッカー」との間の位置は、句の区切り位置である。

【0107】一方、「全員ロッカー支給」という発話が行われた場合において、それが、「全員に、ロッカーを支給する」という意味であれば、発話「全員ロッカー支給」における「全員」と「ロッカー」との間には、一般に、ポーズが挿入されたり、また、その間のイントネーションが低く高に変化する。

【0108】従って、この場合、発話「全員ロッカー支給」における「全員」と「ロッカー」との間のポーズの有無やイントネーション等のプリミティブなブロンディ情報を用いることにより、その間の位置が、句の区切り位置かどうかを判定することができる。

【0109】なお、音声認識結果についての句の区切り位置の情報は、機械翻訳部2において用いることにより、その音声認識結果の構文解析等を行う際に、その解析を誤ることを防止することができる。

【0110】ここで、図11に、ブロンディ情報抽出部19が出力するブロンディデータを示す。

【0111】図11(A)においては、ブロンディ情報としての高低アクセント、強弱アクセント、および句

[illegible]

[01119] として、テキスト解析部 31 が出力する記号列は、規則合成部 32 に供給され、規則合成部 32 では、ステップ S8 において、その記号列に基づいて、規則音声合成が行われる。ここで、テキスト解析部 31 が出力する記号列は、プロンディティク付き翻訳結果に含まれていた、テキスト解析部 31 で変換されなかったプロンディティクが含まれており、規則音声合成部 32 は、その記号列に含まれるプロンディティクを、必要に応じて用いて処理を行う。これにより、規則音声合成部 32 では、例えば、入力音声はほぼ同一の発話速度による合成音や、入力音声は発話スピードの性別、あるいは口頭を反映した合成音等のデータを生ずる。このデータは、DA 変換部 33 においてアナログ信号に変換され、これにより、目的言語による翻訳文に対応する合成音が、スピーカ 6 から出力される。また、表示部 4 では、機械翻訳部 2 から供給される目的言語による翻訳文が表示され、処理を終了する。

【0120】以上の処理によれば、精度の高い朝歌を行い、また、朝歌文を、違和感のない合成音で出力することが可能となる。

【0121】即ち、例えば、原音群による発音が、「I  
 ao like there.”であった場合には、音声認識部1のマッ  
 チング部16において、例えば、図12(A)に示すよ  
 うなワードグラフが、音声認識結果として得られ、図  
 12(B)に示すようなブロンディータ19において、例  
 えば、図12(B)に示すようなブロンディータが得られ  
 する。なお、図12(B)の発音の商標では、図1  
 2(B)で説明した場合と同様に、ブロンディ情報のうち  
 の強アクセントが存在する開始位置および終了時刻だけ  
 が記録されたブロンディータを用いている。

[10122] この場合、マツチンズ部1.6では、ワードグラフとプロソディックデータから、図12(C)に示すようなプロソディック付着認識結果が生成される。ここで、図12(C)に示したプロソディック付着認識結果“*I also like <stress>her/<stress>.”は、強アクセントを被すプロソディックの開始タグ(<stress>)と終了タグ(</stress>)の間にある単語“her”に、強アクセントがあることを表している。*

【0123】機械駆動部2には、図12(C)に示したようなプロシディタ付き駆動線路が供給され、モータ部2.1において解析が行われた後、言語変換部2.2へ送られる。

2に供給される。音階交換部22では、ブロードタイタグ  
付を認識結果と、音階交換データバス26における各  
音階交換データとのマッチングが行われ、これにより、  
ブロードタイタグ付を認識結果に最も合致する音階交換デ  
ータが検出される。

【0124】即ち、いま、音節変動部22が参照する音節変動データベース26に、図6に示した変動テーブルが記憶されているとともに、テキスト解析部21によるブロンディグ付けの結果“*I also like <tre so>her</stress>.”の解析結果として、“*NP1 also V <tre so>her</stress>.”が得られたとする。**

【0125】この場合、プロシデチャ付を既読結果1  
also like <stream>her</stream>。の解析結果“NP1 al  
also vi <stream>NP2 </stream>”に最も合致する音節変換  
データは、図6 (D) の変換テーブルの1行目に記述さ  
れて、この“NP1 also vi NP2”であり、音節変換2 2 2  
では、この音節変換データが抽出される。即ち、解析結  
果の最初の世居“NP1”と音節変換データの最初の世居“NP  
1”、解析結果の2番目の世居“also”と音節変換データの

2番目の単語"also"、解析結果の3番目の単語"v"と音韻変換データの3番目の単語"v"、および解析結果の4番目の単語"NP2"と音韻変換データの4番目の単語"NP2"は、いずれも一致しており、解析結果の4番目の単語"NP2"を囲むブロンディクタグ<stress>および<stress>と、音韻変換データの4番目の単語"NP2"を囲むブロンディクタグ"NP2"は、いずれも一致する。従って、音韻変換結果2.2では、図6、(D)の変換テーブルの1行目に記述されているNP1 also v NP2e、が、ブロンディクタグ付き照写結果"1 also like <stress>her/<stress>"の解析結果"NP1 also v <stress>NP2e/<stress>"に最も一致する音韻変換データとして検出される。

【0126】そして、音節交換部22は、抽出した音節交換データ「N1 also v・NP2」と、その対応「NP1はNP2もv。」を、ブロンディタグ付き音声認識結果「I also like <stress>her</stress>」とともに、テキスト生成部23に出力する。

【10127】なお、ブロンディタグでも陰陽結果のブロンディタグを無視した場合、即ち、ブロンディタグのない音声陰陽結果を解析して、その解析結果に最も合致する音節要素を抽出する場合には、図6 (D) の“*pi alao v a n p 2 a*”と、“*n p 2 i a l a o v n p 2 a*”の両方が検出されることとなり、その優先をつけることが困難であるため、以降の処理を統行する際の前提となる。

【0128】また、プロシディタリ付与認識結果と音行変換データとのマッチングの方法としては、一般の文章検索で用いられている、例えば、論理式による検索方法や、ベクトル空間法による検索方法等も採用することができ、論理式による検索方法を採用した場合には、プロシディタリ付与認識結果を構成する語句を用いて、論

理式（例えば、キーワードとなる語句を、ANDや、OR、NOT等を用いて接続したもの）が作成され、その論理式と言語変換データとは比較することにより、プロシディック語句を認識結果と最も適合する言語変換データが検出される。また、ベクトル空間法による検算方法が検算される場合には、プロシディック語句を構成する語句の中から、クエリベクトル(Query Vector)が作成されるとともに、言語変換データに基づいてベクトルが作成され、その2つのベクトルがなす角度を最も小さくする言語変換データが、プロシディック語句を認識結果に最も適合する言語変換データとして検算される。

【0129】テキスト生成部23は、音節変換データ「N1 also V NP2a」、その対応「NP1はNP2も」、およびプロンティダグ付き音節変換結果「I also like (strengthen/attends)」を受信すると、音節変換データにおけるNP1に対応する「I」に対応する「like」、NP2に対応する「her」を、それぞれ目的語である日本語に換算し、これにより、NP1に対して「私」、Vに対して「好きです」、NP2に対して「彼女」を、それぞれ得る。そして、テキスト生成部23は、それらの日本語を、対応「NP1はNP2もV」にあてはめ、翻訳結果として、図1の2(D)に示すように、「私は彼女も好きです。」を得る。

【0130】この翻訳結果「私は彼女も好きです。」は、音声合成部3に供給され、これにより、「I also like her.」に対する翻訳結果として、合成音「私は彼女も好きです。」が出力される。

【0131】なお、図2の音声認識システムでは、例えば、前述の特開平6-332494号公報に記載されているように、原音語による入力音声においてアクセントのある語句に対応する目的語語の語句にアクセントを付して出力することも可能である。

[0 1.3.2] 即ち、例えば、上述の入力音声“i also li  
ke shwa”については、そのプロソディタグ付き転写結果  
果“i also like <stress>her/<stress>”から、“her”に  
アクセントを付加することを出発することができる。さら  
に、その機械転写時に、アクセントのある“her”に対応  
する日本語段が「彼女」であることを出発することので  
きる。従って、転写結果「私は彼女が好きです。」の  
「彼女」に、強弱を表すプロソディタグを付加し、プロ  
ソディタグ付き転写結果を、音声合成部3に出力すること  
により、原音声による入力音声に対してアクセントの  
ある語句に目的言語の語句に対応してアクセント  
を付した合成音を出力することができる。

【0133】但し、英圏におけるアクセントは、強勢アクセントであるのに対して、日本圏におけるアクセントは、高低アクセントであるから、取附結果に付加するブロッディティグは、ブロッディティグ付を配属結果“also like (a)tresnber/(a)tresn.”に含まれる強勢アクセントを表すブロッディティグではなく、高低アクセントを表す

サブゾンディタグとするのが好ましい。いま、高低アクセントを表すゾンディタグを、図 11 (C) に示したように、<hi-pleth>および<hi-pleth>で表すと、ゾンディタグ付を翻訳結果は、図 12 (E) に示すように、「私は<hi-pleth>彼女<hi-pleth>も好きです。」となる。

【10134】以上のように、機軸(翻訳第2)において、原書語を目的書語に翻訳するにあたり、原書語が、原書語または目的書語のうちの少なくとも一方のブロンディ情報とともに記述されている英語テーブルに基づいて、翻訳を行うようにしたので、精度の高い翻訳を行い、また、翻訳文を、違和感のない合成音響出力することが可能となる(翻訳結果としてよき悪い合成音響出力することが可能となる)。

【0136】即ち、例えば、図13に示すように、英語のユーザが、「I also like there.」と発話した場合に、は、図14に示すように、「私は彼女も好きです。」という正しい翻訳結果の合成音が出力される。なお、図13においては（次に説明する図16においても同様）、英語する言語には、下線を付してある。

【0136】また、例えば、図16に示すように、英語の「Henry has arrived.」と英語した場合には、「ヘンリーが到着しました。」という正しい翻訳結果（英語の英語のアクセントから考えて、適切な助動詞である「が」を用いた翻訳結果）の合成音が、正しい位置にアクセントが付与されて出力される。

【0137】さらに、例えば、図16に示すように、英語のニューズが、“Henry has arrived.”と英語した場合に、**は**、**ヘンリー**は頭等しました。」という正しい翻訳結果（英語の英語にアクセントがないことから考えて、適切な助詞である**「は」**を用いた翻訳結果）の合成音が、正しい位置にアクセントが付与されて出力される。

【0138】なお、例えば、“English teacher”を正しく「イギリス人の先生」に翻訳することにも、“English teacher”を正しく「英語の先生」に翻訳することは、音声認識部1（図4）における辞書データベース1777に、“English teachers”と“English teacher”とを、別の単語（語句）として登録した辞書を登録しておおき、“English teachers”と“English teacher”の音声認識結果を区別して出力するようにすることにも、機械的翻訳部2においても、“English teachers”と“English teacher”を区別して取り扱うことで行われることが可能である。

【0.139】しかしながら、同一の単語内ではあるが、アクセントのある単語と異なるアクセントを有する単語とを区別して音声認識を行うための略語を作成するには、例えば、複合語化することによりアクセントのある単語が移動する単語列で、かつその略記結果が、アクセントのある単語が移動しない場合と異なるものを調査して、略語に反映させる必要がある。従って、そのような略語の作成にあたっては、

ては、音声認識部1を製作する製作者に、音声認識処理のための知識以外の言語学的な知識等も要求されるため、その作成は、現実には困難であると考えられる。

[illegible]

[0141] 具体的には、例えば、図16と図16とに示したように、同一単語列で構成される発話“Henry has arrived.”と“Henry has arrive.”との区別分けを行なう場合において、それぞれの発話のブロンディ情報のうち、唇閉結果に影響を与えるのは、“Henry”にアクセントがあるかどうかという点だけである。

【0142】従って、常時、全音声区間に亘ってブロードディンク情報を抽出することは、音声認識部1の後に処理を行う燃焼制御部2や音声合成部3において用いられないブロードディンク情報を抽出することがあり、処理効率が良いとはいえない。

【0143】そこで、音声情報部1（図4）のブロードバンド情報抽出部19は、その後段で処理を行う機械的取除2や音声合成部3から要求があった場合にのみ、必要となる音声区間だけを抽出に、必要なブロードバンド情報を抽出させるようにすることができ、この場合、処理効率を向上させることができる（処理量を少なくする）ことができる。

【0144】なお、同様の観点から、機械的取除部2に、その後段で処理を行う音声合成部3から要求があった場合にのみ所定の処理を行わせるようにして、機械的取除部2における処理効率を向上させることが可能であ

【0145】この場合、音声認識部1、機械翻訳部2、音声合成部3では、相互に、図17に示すようなやりとりが行われる。なお、図17において、縦方向は、時間の経過を示す。

【0146】図17に示した場合においては、音声が入力され、音声認識部1において音声認識が行われ、その音声認識結果D<sub>1</sub>が、機械翻訳部2に出される。このとき音声認識部3のブロンディ情報抽出部19では、ブロンディ情報の抽出は行われて、従って、音声認識部1は、ブロンディデータを含んでいない音声認識結果D<sub>2</sub>を出力する。なお、ここでは、ブロンディ情報抽出部19において、一切のブロンディ情報を抽出しないこととしたが、機械翻訳部2や音声合成部3において処

理に用いられる頻度の高いプロツディ情報だけは常時抽  
出するようにし、音声認識結果D<sub>1</sub>に含めるようにする  
ことが可能である。

【0147】 燃気取配器2は、音声認識部1から音声認識結果D<sub>1</sub>を受信すると、その音声認識結果D<sub>1</sub>を用いて処理を行い、その処理中に、ブロンディ情報が必要となった場合には、処理を中断して、必要なブロンディ情報を要求するリクエスト信号D<sub>2</sub>を、音声認識部1に出力する。

【0148】ここで、リクエスト番号D<sub>2</sub>によって要求されるプロパティ情報としては、例えば、ある単語に対応する音声に、強調アクセントがしてある単語とそうでない単語とを区別して、強調アクセントがつけられているかどうか、付されていないかどうか、どの程度の割合でアクセントがあるか）や、ニューズ（知語）の性別は男性または女性のものであるか、ある単語とその直後の単語との間に音声上の区別があるかどうか、文末のイントネーションは上り調子なのか下り調子なのによって要求された情報がある。

【0149】 音声認識部1のブロンディ情報抽出部19は、機械翻訳部2から、リクエスト番号D<sub>2</sub>を受信すると、そのリクエスト番号D<sub>2</sub>によって要求されているブロンディ情報を、音声データベース14Aおよび特徴量データベース14Bを参照することで求め、リクエスト番号D<sub>2</sub>に対する応答D<sub>3</sub>として、機械翻訳部2に出力す

【0150】根據圖面2は、応答D<sub>3</sub>を受信すると、その応答D<sub>3</sub>に含まれるプロンディ情報を用いて処理を遂行し、以下、処理中、プロンディ情報が必要となった場合には、処理を中断して、そのプロンディ情報を要求するリクエスト信号を、音声認識部11に出力し、音声認識部1から、必要なプロンディ情報としての応答を待て、処理を再開することと繰り返す。

〔0161〕そして、機械翻訳部2は、音声認識部1からの音声認識結果D<sub>1</sub>の翻訳を完了すると、その翻訳結果D<sub>4</sub>を、音声合成部3に出力する。

【0152】音声合成部2は、機械翻訳部2から翻訳取得した単語D<sub>1</sub>を用いて処理を行う。単語D<sub>1</sub>が受信すると、その翻訳結果D<sub>2</sub>を用いて処理を行う。この処理中に、プロテクト情報や、その他の情報が必要となった場合には、処理を中断して、必要なプロテクト情報やその他の情報を要求するリクエスト信号D<sub>6</sub>を機械翻訳部2に出力し、あるいは同様のリクエスト信号D<sub>7</sub>を音声認識部1に出力する。

【0153】ここで、リクエスト信号D<sub>6</sub>によって、機  
械制御部2に要求する情報としては、例えば、同教習  
音の発音（例えば、“read”は、現在形と過去形とで発  
音が異なるが、その発音）、同教習アクセント語の  
アクセント（例えば、“increase”は、品詞によってアクセ  
ントが異なるが、そのアクセント位置）、複合語化し  
ている可能性のある単語別のイントネーション（例え  
ば、“English teacher”のイントネーション） 文来

インターネット・セッションは上り関子または下り関子のいずれでも、あるか時といった情報がある。なお、これらの情報は、常時、雑誌編集課2が出力する翻訳結果としてののテクニカルに、プロンディタクと同様のタグを用いて含めるようにすることも可能である。

【0164】また、リクエスト番号Dによって、音声認識部11に要求する情報としては、入力音声の発話速度や、ピッチ、ユーザの性別、感情、口調、文長のイントネーションは上り調子または下り調子のいずれであるか等といった情報がある。

【0165】送信機部22は、音声合成部3から、リクエエスト信号D<sub>0</sub>を受信すると、そのリクエエスト信号D<sub>0</sub>に基づいて要求される情報を、上述のように、テキスト解析部21や、音韻変換部22、テキスト生成部23が処理中に保持しておいた情報を参照することによって、リクエエスト信号D<sub>0</sub>に対する応答D<sub>0</sub>として、音声合成部3に出力する。

【0166】また、音声認識部1は、音声合成部3から、リクエスト信号D<sub>7</sub>を受信すると、そのリクエスト信号D<sub>2</sub>によって要求されている、例えばブロンディ情報や、「音声データバッファ14Aおよび特許量バッファ14Bを参照することや、リクエスト信号D<sub>7</sub>に対する応答D<sub>8</sub>として、音声合成部3に出力する。

【01671】音合成部3は、応答D<sub>6</sub>やD<sub>8</sub>を受信すると、その応答D<sub>6</sub>やD<sub>8</sub>に含まれる情報を用いて処理を実行し、以下、処理中に、プロシディ情報、その他の必要な情報が必須な場合とは、処理を中断して、必要な情報を要求するリクエスト信号D<sub>6</sub>やD<sub>7</sub>を、機械翻訳部2やプロシディ情報部1に出だし、機械翻訳部2やプロシディ情報部1から、必要な情報としての応答を得て、処理を再開することゝなり得る。

【0158】なお、音声合成部3から、音声認識部1に対しては、直接、リクエスト番号を供給するのではなく、機械翻訳部2を介して、間接的に、リクエスト番号を供給するようにし、また、音声認識部1から、音声合成部3に対しては、直接、応答を供給するのではなく、機械翻訳部2を介して、間接的に、応答を供給するようにすることが可能である。

[0159] 即ち、例えば、入力音声「I also like this」に対して、その日本語原稿「私は彼女も好きです。」の合成音を出力する場合には、「私は\*彼女も好きです。」のように、入力音声中強調されている単語「her」に対応する「彼女」を強調するときには、音声合成部3からは、機械翻訳部2から、「私は彼女も好きです。」という翻訳文を受け取った後に、その中の「彼女」に対応する入力音声にアクセントがあるかどうかの調査を要求するリクエスト番号を、機械翻訳部2に出力する。機械翻訳部2は、そのリクエスト番号による要求に關して、「彼女」に対応する入力音声における単語が「her」であることを認識しているため、その「her」に対応する入力

20

声にアクセントがあるかどうかの調査を要求するリクエスト番号を、音声認識部1に出力する。音声認識部1では、入力音声「I also like hers」における「her」に、アクセントがあるかどうかの調査を、いまの場合、「her」にアクセントがあるため、その旨の応答が、リクエスト番号を出力してある機械翻訳部2に供給される。機械翻訳部2は、音声認識部1から、「her」にアクセントがある旨の応答を受信すると、「彼女」に対応する入力音声にアクセントがある旨の応答を、リクエスト番号を出力してある音声合成部3に出力する。音声合成部3は、このようにして、機械翻訳部2から、「彼女」に対応する入力音声にアクセントがある旨の応答を受信すると、翻訳文「私は彼女も好きです。」における「彼女」を強調した合成音「私は\*彼女も好きです。」を生成して出力する。

【0160】次に、図118のフローチャートを参照して、機体超群2や音声合成部3がリクエスト信号を出して、機体超群2や音声認識部1や機体超群2が、そのリクエスト信号に対する応答を出力する場合の、音声認識部1の動作について説明する。

【0161】音声認識部1では、ステップS11において、音声が入力され、またはリクエスト信号が、機体超群2や音声合成部3から送信されてきたかといったイベントが生じるまで、待ち時間がおかれ、イベントが生じると、ステップS12に進み、どのようなイベントが生じたかが判定される。ステップS12において、音声が入力されたか、その入力された音声で認識され、ステップS13に進み、ステップS14では、ステップS13における音声認識結果が、機体超群2に出力され、ステップS11に戻り、以下、同様の処理を繰り返す。なお、ステップS14においては、音声認識結果を、機体超群2に出力するとともに、必要に応じて、表示部5に供給して表示させるようにすることも可能である。

【0162】一方、ステップS112において、リクエスト番号が送信されてきたというイベントが生じたこと判定された場合、ステップS116に進み、そのリクエスト番号に対応して、ブロードキャスト情報が抽出される。すなわち、ステップS115では、ブロードキャスト情報抽出部198において、例えば、アクセンとの抽出や、発話速度の計算、ユーザが男性または女性が行われるかとの識別、非言語的な音情報の抽出などが行われる。そして、ステップS116に進み、抽出されたブロードキャスト情報を表すブロードキャストIDに当たり、リクエスト番号に宛する応答として、そのリクエスト番号を送信してきたブロック（ここでは、接続プログラム2または音声合成部3のうちのいずれか）に供給され、ステップS111に戻り、以下、同様の処理が繰り返される。

【0163】次に、図18のフローチャートを参照し

て、機械翻訳部2や音声合成部3がリクエスト信号を出  
力するとともに、音声認識部1や機械翻訳部2が、その  
リクエスト信号に対する応答を出力する場合の、機械翻  
訳部2の動作について説明する。

[0164] 機械翻訳部2では、ステップS21におい  
て、音声認識部1から音声認識結果が送信されてきた  
か、または音声合成部3からリクエスト信号が送信さ  
れてきたかといったイベントが生じるまで、待ち時間が  
おくれ、イベントが生じると、ステップS22に進み、  
どのようなイベントが生じたかが判定される。ステッ  
プS22において、音声認識結果が送信されてきたとい  
うイベントが生じたとき、ステップS23において、機  
械翻訳部2の動作が、音声認識結果が、テキスト生成  
部による音声認識結果に基づいて、音声認識部22か  
ら、音声認識部22では、ステップS24において、テ  
キスト生成部23に出力される。テキスト生成部  
23では、ステップS25において、音声認識部22か  
らの音声認識結果に基づいて、原言語による音声認識  
結果が、目的言語のテキストに翻訳され、ステップS26に  
進み、その翻訳結果が、音声合成部3に出力される。そ  
して、ステップS27に戻り、同様の処理を繰り返す。

[0166] なお、ステップS26においては、翻訳結  
果を、音声合成部3に出力するとともに、必要に応じ  
て、表示部6に供給して表示させるようにすることも可  
能である。また、テキスト生成部23は、後述するように、そ  
の処理中に、音声認識部1に対して求められる情報(こ  
こでは、プロンディ情報)が必要となる場合は、処理  
を中断し、その情報を要求するリクエスト信号を、音声  
認識部1に送信する。そして、そのリクエスト信号に対  
応して、音声認識部1から情報が送信されてくるのを待  
って、処理を再開する。

[0166] 一方、ステップS22において、音声合成  
部3から、リクエスト信号が送信されてきたというイベ  
ントが生じたとき、ステップS27に進  
み、そのリクエスト信号によって要求されている情報が  
求められる。即ち、上述のように、機械翻訳部2の  
処理中に参照した、辞書データベース24、解析用文法  
データベース25、音声認識データベース26、辞書デ  
ータベース27、および生成用文法データベース28の  
情報を保持しており、その保持している情報を参照する  
ことで、例えば、入力音声やその翻訳結果におけるアク  
セントの位置、単語の発音、非言語的な情報等の、リク  
エスト信号によって要求される情報が求められる。リ  
クエスト信号に対する応答として、そのリクエスト信号  
を送信してきた音声合成部3に供給され、ステップS2

[0175] また、ステップS42において、何らかの  
プロンディ情報が必要であると判定された場合、ステッ  
プS43、S44に順次進み、音質変換部22では、図  
20(A)のステップS33、S34における場合とそ  
れぞれ同様の処理が行われ、これにより、必要なプロン  
ディ情報が、音声認識部1から取得される。そして、ス  
テップS45に進み、そのプロンディ情報を用いて、ス  
テップS41で読み込んだ解析結果が音質変換処理され  
る。

[0176] ステップS45の処理後は、ステップS4  
6に進み、テキスト生成部21が出力する解析結果のす  
べての音質変換を終了したかどうか判定される。ステ  
ップS46において、解析結果のすべての音質変換を、  
まだ終了していないと判定された場合、ステップS41  
に戻り、まだ音質変換していない解析結果を読み込  
み、以下、同様の処理が繰り返される。

[0177] 一方、ステップS46において、解析結果  
のすべての音質変換を終了したと判定された場合、リク  
エーンする。

[0178] 次に、図20(C)のプロチャートを参  
照して、機械翻訳部2のテキスト生成部23が行うステ  
ップS25の処理の詳細について説明する。

[0179] テキスト生成部23は、ステップS61に  
おいて、音質変換部22からの音質変換結果の全部また  
は一部を読み込み、ステップS62に進み、その音質変  
換結果を処理するの、何らかのプロンディ情報が必要  
かどうかを判定する。ステップS62において、特に、  
プロンディ情報が必要でないとき判定された場合、ステッ  
プS63およびS64をスキップして、ステップS65  
に進み、ステップS51で読み込んだ音質変換結果を対  
象に、目的言語のテキストが生成される。

[0180] また、ステップS62において、何らかの  
プロンディ情報が必要であると判定された場合、ステッ  
プS63、S64に順次進み、テキスト生成部23で  
は、図20(A)のステップS33、S34における場  
合とそれぞれ同様の処理が行われ、これにより、必要な  
プロンディ情報が、音声認識部1から取得される。そし  
て、ステップS65に進み、そのプロンディ情報を用い  
て、ステップS61で読み込んだ音質変換結果を対象  
に、目的言語のテキストが生成される。

[0181] 音質変換部22の処理後は、ステップS6  
6に進み、音質変換部22が出力する音質変換結果のす  
べてについて、テキストの生成を終了したかどうか判  
定される。ステップS66において、音質変換結果のす  
べてについてのテキストの生成を、まだ終了していない  
と判定された場合、ステップS61に戻り、まだテキス  
トを生成していない音質変換結果を読み込み、以下、  
同様の処理が繰り返される。

[0182] 一方、ステップS66において、音質変換  
結果のすべてについてのテキストの生成を終了したと判

定された場合、リターンする。

[0183] 次に、図21のプロチャートを参照し  
て、機械翻訳部2や音声合成部3がリクエスト信号を出  
力するとともに、音声認識部1や機械翻訳部2が、その  
リクエスト信号に対する応答を出力する場合の、音声合  
成部3の動作について説明する。

[0184] 音声合成部3では、ステップS61におい  
て、機械翻訳部2からの翻訳結果の送信というイベン  
トが生じるまで、待ち時間がおくれ、そのイベントが生  
じると、ステップS62に進み、機械翻訳部2からの翻訳  
結果が、テキスト生成部31において解析され、その解  
析結果が、規則合成部32に供給される。規則合成部3  
2では、ステップS63において、テキスト生成部31  
が出力する解析結果に基づいて、規則音声合成が行わ  
れ、合成音のディジタルデータが生成される。このディ  
ジタルデータは、DA変換部33に供給され、DA変換  
部33では、ステップS64において、規則音声合成部  
32からのディジタルデータがD/A変換され、これに  
より、翻訳結果に対応する合成音が、スピーカ5から出  
力される。そして、ステップS61に戻り、以下、同様の  
処理を繰り返す。

[0185] なお、テキスト生成部31、規則合成部3  
2、およびDA変換部33は、その処理中に、音声認識  
部1や機械翻訳部2において求められる情報が必要とな  
った場合は、処理を中断し、その情報を要求するリクエ  
スト信号を、音声認識部1や機械翻訳部2に送信する。  
そして、そのリクエスト信号に対応して、音声認識部1  
や機械翻訳部2から情報が送信されてくるのを待つて、  
処理を再開する。

[0186] 次に、図22を参照して、図21のステッ  
プS62、S63の処理それぞれの詳細について説明す  
る。

[0187] まず最初に、図22(A)のプロチャー  
トを参照して、音声合成部3のテキスト生成部31が行  
うステップS62の処理の詳細について説明する。

[0188] テキスト生成部31は、ステップS71に  
おいて、機械翻訳部2からの入力、即ち、翻訳結果の全  
部または一部を読み込み、ステップS72に進み、その  
部または一部を処理するの、何らかのプロンディ情報(さ  
らには、その部の情報)が必要かどうかを判定する。ス  
テップS72において、特に、プロンディ情報が必要で  
ないと判定された場合、ステップS73およびS74を  
スキップして、ステップS75に進み、翻訳結果の解析  
が行われる。

[0189] また、ステップS72において、何らかの  
プロンディ情報が必要であると判定された場合、ステッ  
プS73に進み、テキスト生成部31は、その必要な  
プロンディ情報を要求するリクエスト信号(その必要な  
の、どのような種類のプロンディ情報が必要なのかを告  
ぐ番号)を、音声認識部1または機械翻訳部2に出力す

場合、リターンする。

[0197] 以上のように、音声認識部1から、リクエスト番号に対応して送信されてくる語句の音声認識結果の解析が行われ、さらに、その解析結果に基づいて、適切な音響変換データが抽出される。そして、機械認識部2は、その音響変換データを用いて、音声認識結果の音響変換(細訳)を行うのに必要な語句を求め、その語句を要求するリクエスト番号を、音声認識部1に送信する。

[0198] なお、本実施の形態では、音声認識部1から機械認識部2に対して、音声認識結果を、常に送信するようにしたが、音声認識部1から機械認識部2に対しては、機械認識部2から必要な音声認識結果を要求するリクエスト番号を送信されてきた場合のみ、その必要な音声認識結果を送信するようにすることが可能である。

[0199] 即ち、例えば、機械認識部2では、音響変換部22において、音声認識結果と音響変換データとのマッチングが行われ、これにより音声認識結果に最も合致する音響変換データが抽出されるが、このマッチングにおいては、英語に含まれる一部の語句が必要となしな場合がある。具体的には、例えば、音響変換データが、述語(動詞など)ごとに分類されている場合においては(例えば、「OOを下し」や「OOを食べます」など)、マッチングにおいて最初に必要となる情報は、英語中の述語と、その述語に係る語句であり、また、例えは、音響変換データが動詞(例えば、「OOを××」や、「OOを××」など)ごとに分類されている場合には、マッチングにおいて最初に必要となる情報は、英語中の動詞と、その動詞に係る語句が必要となる。つまり、英語に含まれる一部の語句が必要となる場合であり、英語に含まれる一部の語句が必要でない場合には、英語の内の一部の語句と音響変換データとのマッチングには、上述した論理式やベクトル空間などの手法を用いることができる。

[0200] この場合、音声認識部1から機械認識部2に対しては、最初に、英語に含まれる一部の語句の音声認識結果だけを供給し、使用する音響変換データが抽出されてから、英語全体の音声認識結果を供給すれば十分である。

[0201] そこで、音声認識部1および機械認識部2の間では、音声認識結果について、図17で説明した場合と同様のやりとりを行わせることが可能である。

[0202] 即ち、この場合、音声認識部1においては、英語が音声認識され、その認識結果が得られると、その旨を示すイベントが、機械認識部2に送信される。機械認識部2は、音声認識部1から、音声認識結果が得られた旨のイベントを受信すると、その音声認識結果を処理するの、最初に必要となる語句を要求するリクエスト番号を、音声認識部1に送信する。音声認識部1は、機械認識部2からリクエスト番号を受信すると、そのリクエスト番号によって要求されている語句の音声認識結果を、機械認識部2に送信する。

る。そして、そのリクエスト番号に対応する応答として必要な音響変換データが、音声認識部1または機械認識部2から送信されてくるのを待つ、ステップS74に進む。ステップS74では、音声認識部1または機械認識部2から送信されてくる、リクエスト番号に対応する応答としての必要な音響変換データを受信され、ステップS75に進み、その音響変換データを用いて、細訳結果が解析される。

[0190] ステップS75の処理後は、ステップS76に進み、機械認識部2が出力する細訳結果のすべての解析を終了したかどうか判定される。ステップS76において、細訳結果のすべての解析を、まだ終了していないと判定された場合、ステップS77に戻り、まだ解析していない細訳結果が読み込まれ、以下、同様の処理が繰り返される。

[0191] 一方、ステップS76において、細訳結果のすべての解析を終了したと判定された場合、リターンする。

[0192] 次に、図22(B)のプロローチャート参照して、音声合成部3の規則合成部32が行うステップS63の処理の詳細について説明する。

[0193] 規則合成部32は、ステップS81において、テキスト解析部31からの細訳結果の解析結果の全部または一部を読み込み、ステップS82に進み、その解析結果を処理するの、何らかの音響変換データが必要かどうかを判定する。ステップS82において、特に、音響変換データが必要でない判定された場合、ステップS83およびS84をスキップして、ステップS85に進み、ステップS85で読み込んだ細訳結果に合った規則音声合成が行われる。

[0194] また、ステップS82において、何らかの音響変換データ(さらには、その他の情報)が必要であると判定された場合、ステップS83、S84に進み、規則合成部32では、図22(A)のステップS73、S74における場合とそれと同様の処理が行われ、これにより、必要な音響変換データが、音声認識部1または機械認識部2から取得される。そして、ステップS85に進み、その音響変換データを用いながら、ステップS81で読み込んだ細訳結果にしたがった規則音声合成が行われる。

[0203] 機械認識部2では、音声認識部1から、リクエスト番号に対応して送信されてくる語句の音声認識結果の解析が行われ、さらに、その解析結果に基づいて、適切な音響変換データが抽出される。そして、機械認識部2は、その音響変換データを用いて、音声認識結果の音響変換(細訳)を行うのに必要な語句を求め、その語句を要求するリクエスト番号を、音声認識部1に送信する。

[0204] 音声認識部1は、このようにして機械認識部2から送信されてくるリクエスト番号によって要求されている語句を、英語の音声認識結果から抽出し、機械認識部2に送信する。そして、機械認識部2では、音声認識部1から送信されてくる音声認識結果の語句を用いて、音声認識結果を目的語に変換した細訳文を生成する。

[0205] 以上のように、音声認識部1から機械認識部2に対して、機械認識部2から要求のあった音声認識結果の語句だけを送信するようにした場合に、処理効率を向上させ、また、音声認識部2の低下を防止することが可能となる。

[0206] 即ち、機械認識部2では、英語に含まれている語句であっても、その語句が、細訳結果に反映されない場合があり、そのような語句が、音声認識部1から機械認識部2に供給されなくなる結果、処理効率が向上する。また、音声認識部1では、音声認識結果の候補が、そのような英語の音声認識結果の候補を、必要に応じて要求するようにすることで、音声認識部1が、その複数の音声認識結果の候補の中から、ある候補を、最終的な音声認識結果として決定して出力することに起因する音声認識部2の低下を防止することが可能となる。

[0207] 次に、上述した一連の処理は、ハードウェアにより行うこともできるし、ソフトウェアにより行うこともできる。一連の処理をソフトウェアにより行う場合には、そのソフトウェアを構成するプログラムが、専用のハードウェアとしての音声認識システムに組み込まれているコンピュータや、汎用のコンピュータ等にインストールされる。

[0208] そこで、図23を参照して、上述した一連の処理を実行するプログラムをコンピュータにインストールし、コンピュータによって実行可能な状態とするために用いられる、そのプログラムが記録されている記録媒体について説明する。

[0209] プログラムは、図23(A)に示すように、コンピュータ101に内蔵されている記録媒体としてのハードディスク102や半導体メモリ103に予め記録しておくことができる。

[0210] あるいはまた、プログラムは、図23(B)に示すように、フロッピーディスク111、CD-ROM(Compact Disc Read Only Memory)112、MO(Magnet

o optical)ディスク113、DVD(Digital Versatile Disc)114、磁気ディスク115、半導体メモリ116などの記録媒体に、一時的あるいは永続的に格納(記録)しておくことができる。このような記録媒体は、いわゆるパッケージソフトウェアとして提供することができる。

[0211] なお、プログラムは、上述したような記録媒体からコンピュータにインストールする他、図23(C)に示すように、ダウンロードサイト121から、デジタル衛星放送用の人工衛星122を介して、コンピュータ101に無線で転送したり、LAN(Local Area Network)、インターネットといったネットワーク131を介して、コンピュータ101に有線で転送し、コンピュータ101において、内蔵するハードディスク102などにインストールすることができる。

[0212] ここで、本明細書において、コンピュータに各種の処理を行わせるためのプログラムを記述する処理ステップは、必ずしもフローチャートとして記載された順序に沿って時系列に処理する必要はなく、並列的あるいは個別に実行される処理(例えば、並列処理あるいはオブジェクトによる処理)も含むものである。

[0213] また、プログラムは、1のコンピュータにより処理されるものであってもよいし、複数のコンピュータによって分散処理されるものであってもよい。さらに、プログラムは、遠方のコンピュータに転送されて実行されるものであってもよい。

[0214] 次に、図24は、図23のコンピュータ101の構成例を示している。

[0215] コンピュータ101は、図24に示すように、CPU(Central Processing Unit)142を内蔵している。CPU142には、バス141を介して、入力出力インターフェース145が接続されており、CPU142は、入力出力インターフェース145を介して、ユーザによって、キーボードやマウス等で構成される入力部147が操作されることにより指令が入力されると、それにしたがって、図23(A)の半導体メモリ103に対応するROM(Read Only Memory)143に格納されているプログラムを実行する。あるいは、また、CPU142は、ハードディスク102に格納されているプログラム、新品122若しくはネットワーク131から転送され、通信部148で受信されてハードディスク102にインストールされたプログラム、またはドライブ149に装着されたフロッピーディスク111、CD-ROM112、MOディスク113、DVD114、若しくは磁気ディスク115から読み出されてハードディスク102にインストールされたプログラムを、RAM(Random Access Memory)144にロードして実行する。そして、CPU142は、その処理結果を、例えば、出力インターフェース145を介して、LCD(Liquid Crystal Display)やスピーカ等で構成される出力部146から、必要に応じて出力させる。

[0216] なお、本実施の形態では、音声認識、その音声認識結果を翻訳するようにしたが、本発明は、キーボード等を操作して入力された文を翻訳する場合にも適用可能である。  
[0217] また、本実施の形態では、翻訳結果を合成音で出力するようにしたが、本発明は、翻訳結果を、テキストで表示する場合にも適用可能である。  
[0218] さらに、本発明は、日英や英日以外の翻訳にも適用可能である。

[0219] [発明の効果] 本発明の翻訳装置および翻訳方法、並びに記録媒体によれば、入力文を、翻訳文に翻訳するための対応関係が、第1または第2の言語のうちの少なくとも一方のプロシディ情報とともに記述されているデータベースに基づいて、入力文が、その入力文に対応する翻訳文に翻訳される。従って、精度の高い翻訳を行い、また、翻訳文を、違和感のない合成音で出力することが可能となる。

[図面の簡単な説明]

[図1] 従来の音声翻訳システムの一例の構成を示すブロック図である。  
[図2] 本発明を適用した音声翻訳システムの一実施の形態の構成例を示すブロック図である。

[図3] 本発明を適用した音声翻訳システムの外観構成例を示す平面図である。

[図4] 音声認識部1の構成例を示すブロック図である。

[図5] 機械翻訳部2の構成例を示すブロック図である。

[図6] 音質変換データを示す図である。  
[図7] 音声合成部3の構成例を示すブロック図である。

[図8] 図2の音声翻訳システムの動作を説明するためのフローチャートである。

[図9] ワードグラフを示す図である。  
[図10] 図9に続く図である。  
[図11] プロシディデータおよびプロシディタグを示す図である。

[図12] 図2の音声翻訳システムの動作を説明するための図である。

[図13] 図3の音声翻訳システムの動作を説明するための平面図である。

[図14] 図3の音声翻訳システムの動作を説明するための平面図である。

[図15] 図3の音声翻訳システムの動作を説明するための平面図である。

[図16] 図3の音声翻訳システムの動作を説明するための平面図である。

[図17] 音声認識部1、機械翻訳部2、および音声合成部3の間でのやりとりを説明するための図である。

[図18] 音声認識部1の動作を説明するためのフローチャートである。

[図19] 機械翻訳部2の動作を説明するためのフローチャートである。

[図20] 図19のステップS23乃至S26それぞれの処理を説明するためのフローチャートである。

[図21] 音声合成部3の動作を説明するためのフローチャートである。

[図22] 図21のステップS62、S63それぞれの処理を説明するためのフローチャートである。

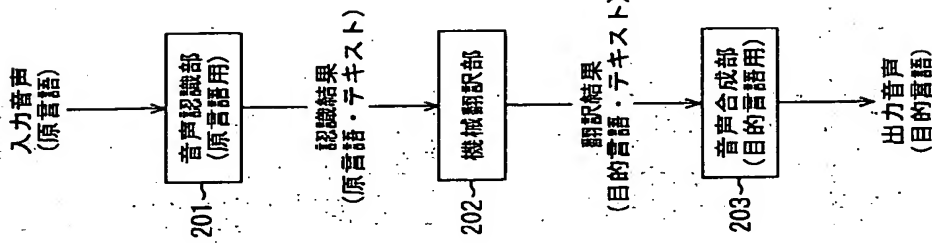
[図23] 本発明を適用した記録媒体を説明するための図である。

[図24] 図23のコンピュータ101の構成例を示すブロック図である。

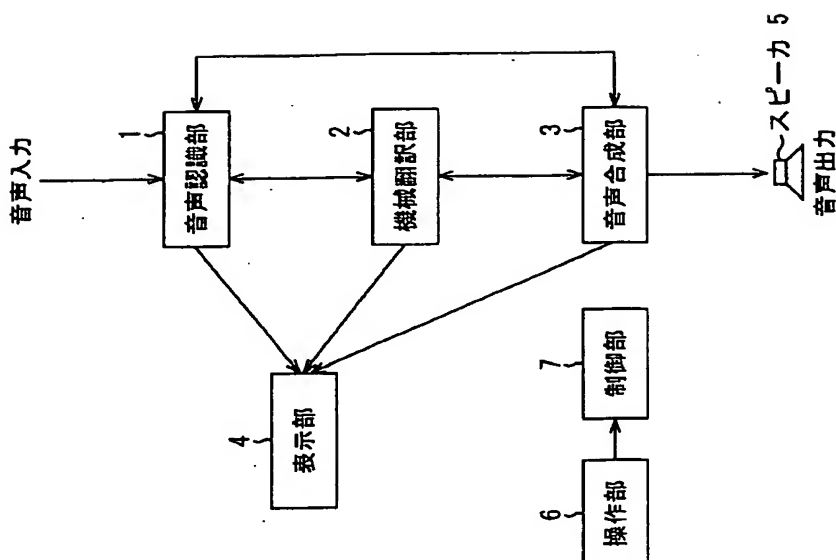
[符号の説明]

- 1 音声認識部, 2 機械翻訳部, 3 音声合成部, 4 表示部, 5 スピーカ, 6 操作部, 7 Aカーソルキー, 8 B決定キー, 9 Cキャンセルキー, 10 制御部, 11 マイク(マイクロフォン), 12 AD変換部, 13 音声抽出部, 14 バッファ部, 15 音声データバッファ, 16 特徴量バッファ, 17 音声データベース, 18 文法データベース, 19 プロシディ情報抽出部, 20 テキスト解析部, 21 音質変換部, 22 テキスト生成部, 23 辞書データベース, 24 音質変換データベース, 25 辞書データベース, 26 音質変換データベース, 27 辞書データベース, 28 生成文法データベース, 29 テキスト解析部, 30 規則合成部, 31 DA変換部, 32 辞書データベース, 33 辞書データベース, 34 辞書データベース, 35 コンピュータ, 36 音質変換データベース, 101 コンピュータ, 102 ハードディスク, 103 半導体メモリ, 111 フロッピーディスク, 112 CD-ROM, 113 MOディスク, 114 DVD, 115 磁気ディスク, 116 半導体メモリ, 121 グラウンドサイト, 122 電源, 131 ネットワーク, 141 バス, 142 CPU, 143 ROM, 144 RAM, 145 入力インターフェース, 146 出力部, 147 入力部, 148 通信部, 149 ドライブ

[図1]



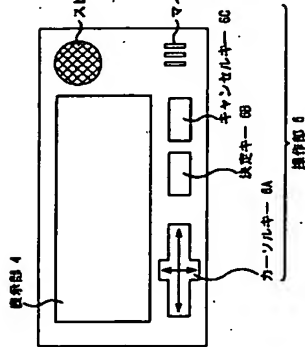
[図2]



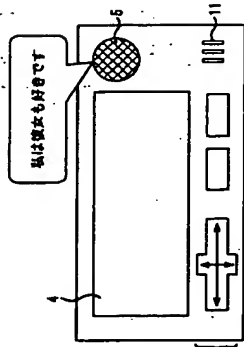
音声翻訳システム

従来の音声翻訳システム

【図 3】



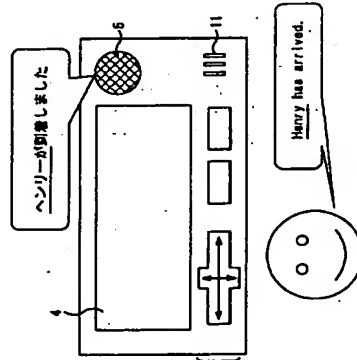
【図 14】



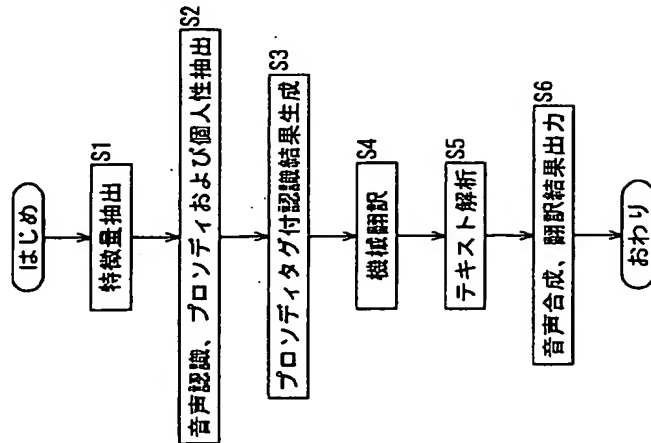
「私は彼女が好きです」という会話をしゃべる

音声認識システムの外観

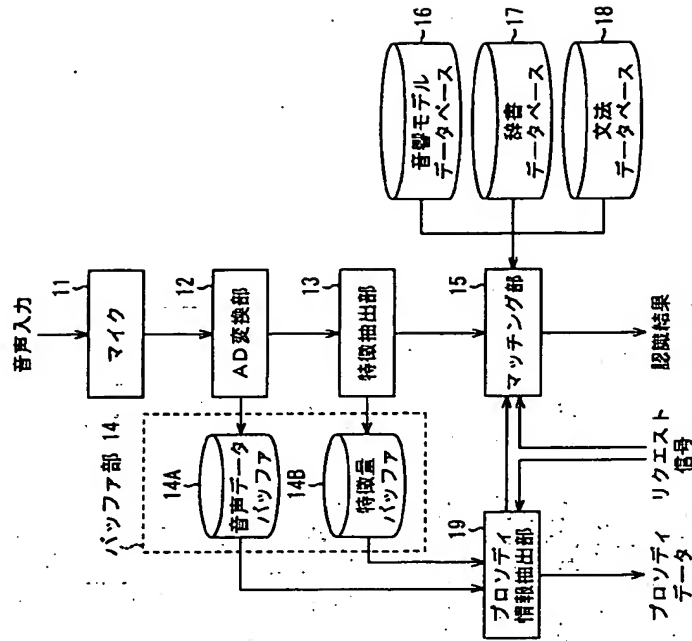
【図 16】



【図 8】

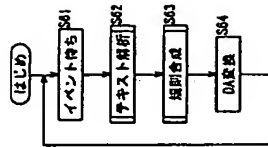


【図 4】



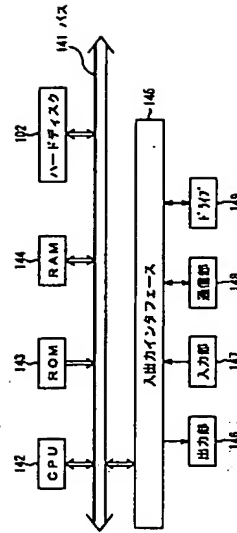
音声合成部3の処理

【図 21】



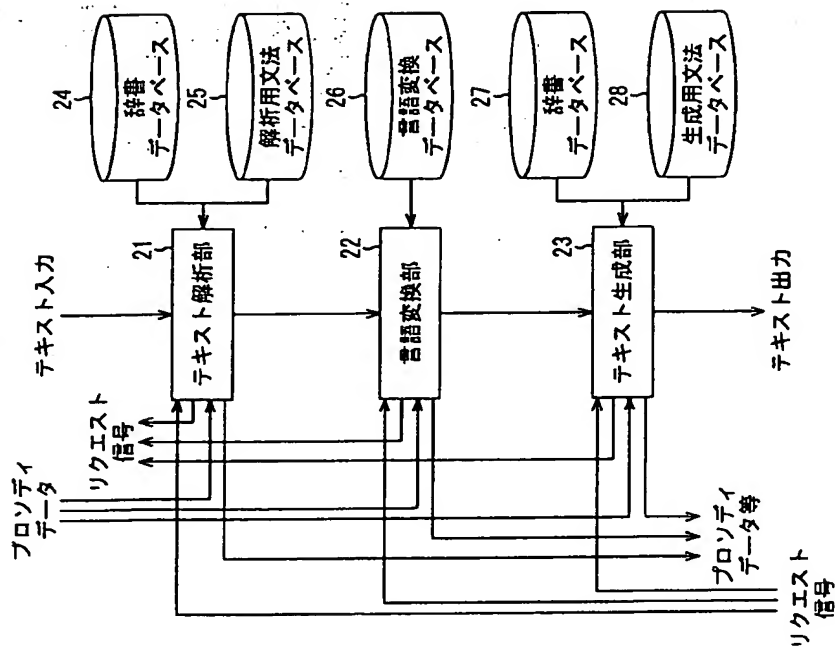
音声認識部 1

【図 24】



コンピュータ 101

【図5】



機械翻訳部 2

【図6】

言語変換データ

原言語	目的言語
English *teacher* *English* teacher	イギリス人の先生 英語の先生

(A)

原言語	目的言語
NP VP *NP* VP	NP は VP NP が VP

(B)

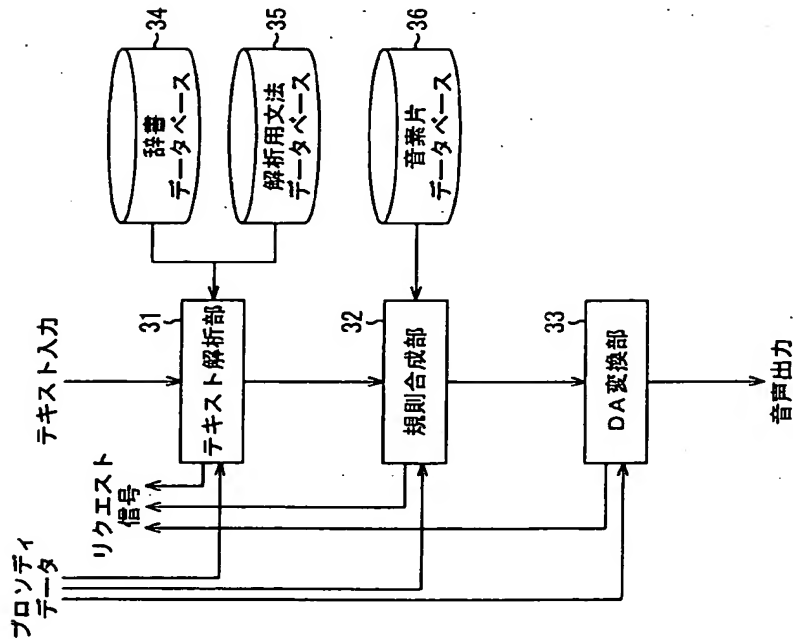
原言語	目的言語
NP は VP *NP* が VP NP が VP	NP VP *NP* VP NP VP

(C)

原言語	目的言語
NP1 also V *NP2*. *NP1* also V NP2.	NP1 は NP2 も V。 NP1 も NP2 を V。

(D)

【図7】



音声合成部 3

【図9】

「この近くにおいていいレストランはありますか」という音声を対するワードグラフ

開始時刻	終了時刻	単語候補	単語スコア
0.00	0.05	(無音)	-666.09
0.00	0.07	(無音)	-754.64
0.00	0.09	(無音)	-866.64
0.00	0.10	(無音)	-1064.94
0.06	0.18	円	-1439.89
0.07	0.18	円	-1268.42
0.07	0.22	そこ	-1694.33
0.09	0.22	ここ	-1402.43
0.09	0.23	ここ	-1507.71
0.09	0.23	明日	-2541.41
0.10	0.22	を	-1808.01
0.18	0.22	の	-440.00
0.22	0.33	の	-1145.82
0.23	0.33	近く	-1052.47
0.33	0.69	に	-3601.02
0.69	0.76	に	-628.41
0.69	0.81	の	-322.48
0.69	0.87	の	-1321.04
0.75	0.81	の	-1817.76
0.75	0.87	の	-668.17
0.76	1.14	の	-1257.99
0.81	1.14	の	-4205.43
0.87	1.14	の	-3440.57
0.87	1.16	の	-2827.15
1.14	1.30	の	-3262.58
1.14	1.35	の	-1706.68
1.14	1.37	の	-2234.33
1.14	1.46	の	-2405.43
1.14	1.58	の	-3380.64
1.14	1.60	の	-4627.71
1.14	1.62	の	-4721.56
1.14	1.63	の	-5028.86
1.14	1.64	の	-5081.19
1.14	1.65	の	-5105.53
1.14	1.65	の	-5282.68
1.16	1.30	の	-1510.35
1.30	1.36	の	-666.54
1.35	1.46	の	-1233.48
1.36	1.46	の	-1134.74
1.37	1.46	の	-1003.09
1.46	1.52	の	-664.54
1.46	1.59	の	-1883.37
1.52	1.59	の	-711.40
1.59	1.61	の	-300.66
1.59	1.64	の	-682.36
1.59	1.62	の	-310.31
1.59	1.70	の	-1083.17
1.59	1.74	の	-1469.65
1.61	1.77	の	-1802.00
1.61	1.74	の	-1301.67
1.62	1.77	の	-1666.44
1.62	1.68	の	-614.30
1.62	1.73	の	-1110.34

【図10】

ワードグラフ(続き)		ワードグラフ(続き)	
1.63	ま	-306.71	
1.64	ま	-307.09	
1.65	は(わ)	-621.59	
1.66	は(わ)	-913.21	
1.67	は(わ)	-1255.66	
1.68	は(わ)	-807.16	
1.69	は(わ)	-1146.61	
1.70	は(わ)	-1904.17	
1.71	は(わ)	-1517.88	
1.72	は(わ)	-3598.02	
1.73	は(わ)	-3394.62	
1.74	は(わ)	-3508.34	
1.75	は(わ)	-3465.62	
1.76	は(わ)	-872.66	
1.77	は(わ)	-405.06	
1.78	は(わ)	-748.10	
1.79	は(わ)	-3029.04	
1.80	は(わ)	-430.22	
1.81	は(わ)	-461.76	
1.82	は(わ)	-2217.06	
1.83	は(わ)	-2221.68	
1.84	は(わ)	-2345.70	
1.85	は(わ)	-2322.88	
1.86	は(わ)	-831.99	
1.87	は(わ)	-1158.18	
1.88	は(わ)	-1237.09	
1.89	は(わ)	-1441.17	
1.90	は(わ)	-1756.40	
1.91	は(わ)	-1868.06	
1.92	は(わ)	-1966.43	
1.93	は(わ)	-744.43	
1.94	は(わ)	-956.69	
1.95	は(わ)	-1144.08	
1.96	は(わ)	-1327.71	
1.97	は(わ)	-1557.73	
1.98	は(わ)	-1642.94	
1.99	は(わ)	-1755.60	
2.00	は(わ)	-1852.22	
2.01	は(わ)	-1852.97	
2.02	は(わ)	-836.11	
2.03	は(わ)	-1028.71	
2.04	は(わ)	-1550.36	
2.05	は(わ)	-1734.84	
2.06	は(わ)	-704.88	
2.07	は(わ)	-821.09	
2.08	は(わ)	-733.73	
2.09	は(わ)	-876.72	
2.10	は(わ)	-3651.57	
2.11	は(わ)	-1881.52	
2.12	は(わ)	-904.61	
2.13	は(わ)	-3306.17	
2.14	は(わ)	-2537.68	
2.15	は(わ)	-1742.01	
2.16	は(わ)	-282.62	

【図11】

(A) プロソディデータ		意味	
開始時刻	終了時刻	高アクセント:6	
0.00	0.05	高アクセント:2	
0.06	0.33	高アクセント:1	
0.81	1.14	高アクセント:2	
0.09	0.33	高アクセント:2	
0.15	0.18	高アクセント:2	

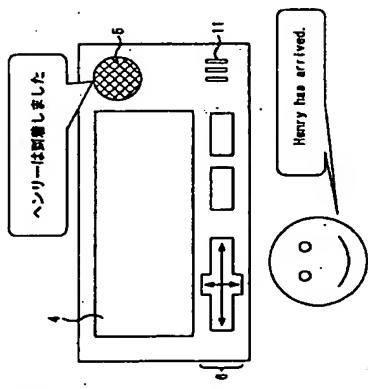
  

(B) プロソディデータ		意味	
開始時刻	終了時刻	高アクセント:6	
0.00	0.05	高アクセント:2	
0.81	1.14	高アクセント:2	

(C) プロソディデータ		意味	
開始時刻	終了時刻	高アクセント:6	
0.00	0.05	高アクセント:2	
0.81	1.14	高アクセント:2	

【図16】



【図12】

(A) ワードグラフ		意味	
開始時刻	終了時刻	高アクセント:6	
0.00	0.05	高アクセント:2	
0.06	0.10	高アクセント:1	
0.07	0.10	高アクセント:2	
0.10	0.20	高アクセント:2	
0.10	0.15	高アクセント:2	
0.15	0.30	高アクセント:2	
0.15	0.20	高アクセント:2	
0.18	0.30	高アクセント:2	
0.20	0.30	高アクセント:2	
0.30	0.39	高アクセント:2	
0.30	0.42	高アクセント:2	
0.39	0.46	高アクセント:2	
0.42	0.46	高アクセント:2	

(B) プロソディデータ(高アクセントのみ)

開始時刻		終了時刻	
0.30	0.37	0.30	0.37

(C) プロソディデータ付き単語認識結果

"I also like<stress> her </stress>"

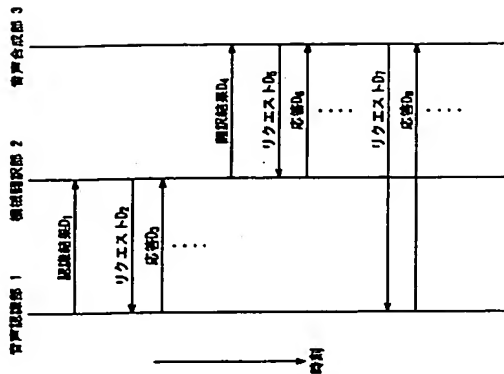
(D) 単語認識結果

"私は 彼女も 好きです"

(E) プロソディデータ付き単語認識結果

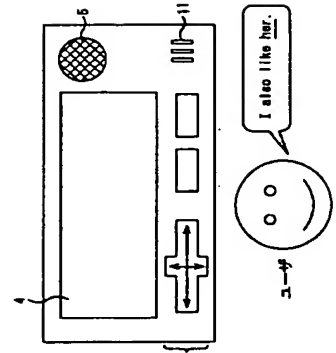
"私は <hi-pitch>彼女</hi-pitch>も 好きです"

【図17】



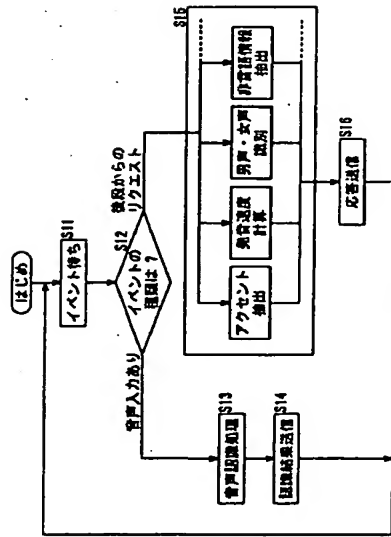
データのやり取り

【図13】



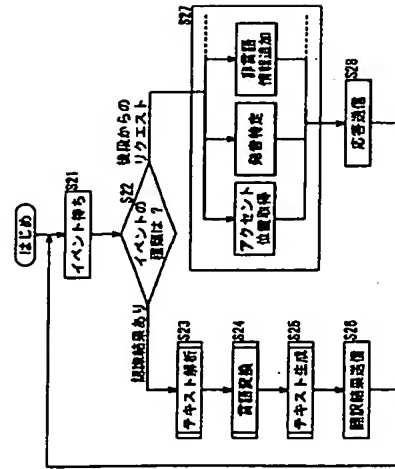
システムに向かって「I also like her」(「her」を強く発音)としゃべった

【図18】



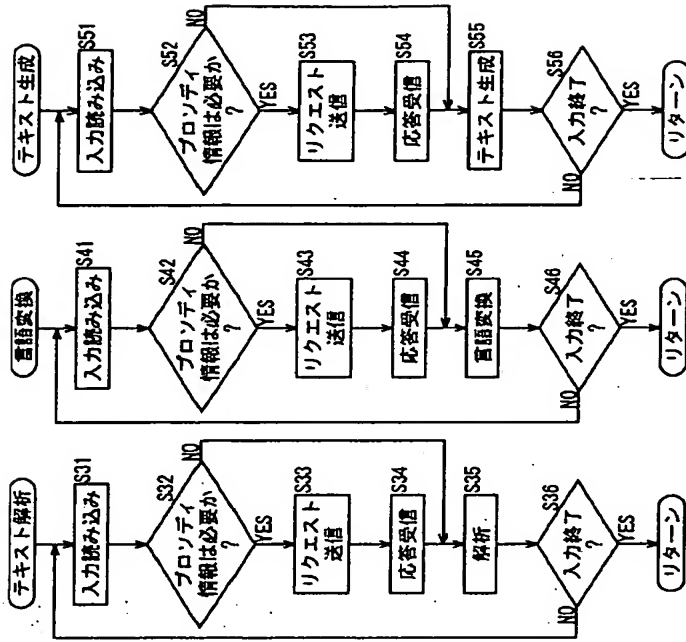
音声認識部1の処理

【図19】



機械翻訳部2の処理

【図20】



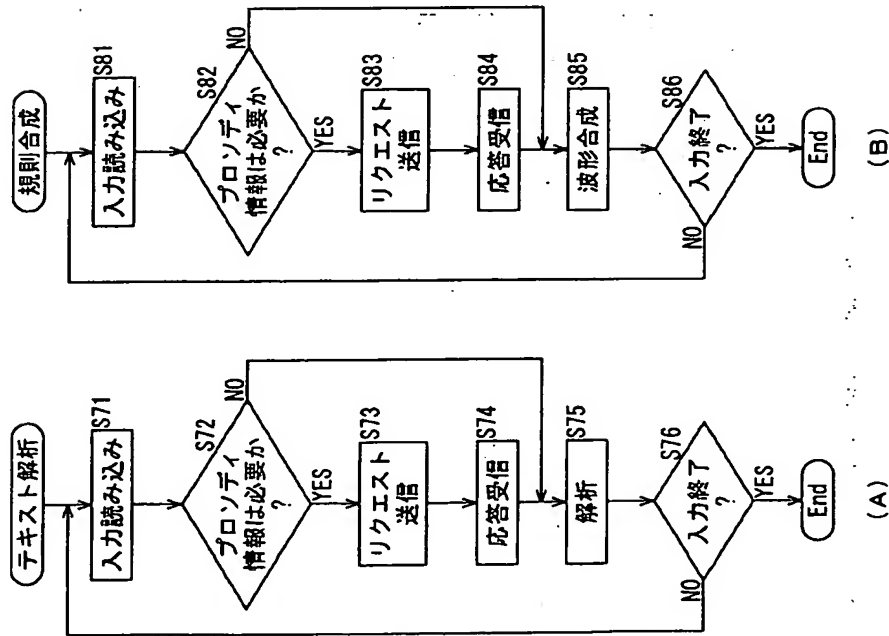
(A)

(B)

(C)

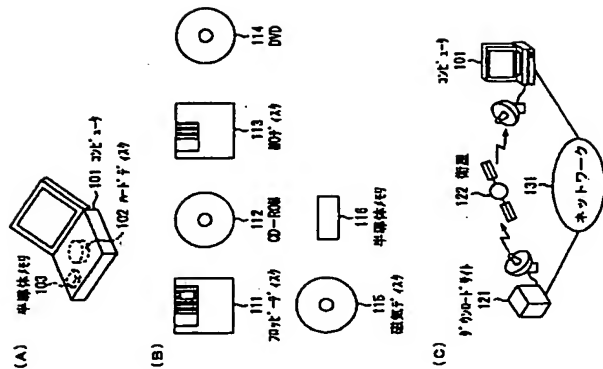
機械翻訳部2の処理

【図22】



音声合成部3の処理

【図23】



フロントページの続き

(61) Int. Cl. <sup>7</sup>	F 1	G 10 L 3/00	6 5 1 C	フーワード (参考)
(72) 発明者	岸 秀樹	Fターム (参考)	6B091 AA06 AA15 BA03 CA21	
	東京都品川区北品川6丁目7番35号	ソニ	CB12 CB32 CC01 EA00	
	株式会社ソニ		ED015 CC13 CC14 KK02 KK04	
(72) 発明者	渡野 廣祐		ED045 AA07 AB03	
	東京都品川区北品川6丁目7番35号	ソニ	9A001 HH14 HZ17 HZ18	
	株式会社ソニ			